UNIVERSITY OF TORONTO
3 1761 01645345 8

F. 2
T

UNIVERSITY OF TORONTO
DEPARTMENT OF PSYCHOLOGY

OCT 21 1919
Nat Comm. Mental Hyg.


Digitized by the Internet Archive
in 2008 with funding from
Microsoft Corporation

http://www.archive.org/details/stanfordrevision00termuoft

THE STANFORD REVISION AND EXTENSION OF THE BINET-SIMON SCALE FOR MEASURING INTELLIGENCE

Educational Psychology Monographs

Edited by Guy Montrose Whipple

No. 18

THE STANFORD REVISION AND EXTENSION OF THE BINET-SIMON SCALE FOR MEASURING INTELLIGENCE

By

Lewis M. Terman, Grace Lyman, George Ordahl, Louise Ellison Ordahl, Neva Galbreath and Wilford Talbert

Assisted by

Herbert E. Knollin, J. H. Williams, H. G. Childs, Helen Trost, Richard Zeidler, Charles Waddle and Irene Cuneo

UNIVERSITY OF TORONTO LIBRARY

BALTIMORE
WARWICK & YORK, Inc.
1917

Psych
T319st

599981
19. 1. 55

Copyright, 1917,
By
Warwick & York, Inc.

Dedicated

by

LEWIS M. TERMAN

to

Those whose loyal coöperation
made the study possible

EDITOR'S PREFACE

The labors of Professor Terman and his co-workers at Stanford University in the critical examination and improvement of the Binet-Simon Scale for Measuring Intelligence are so well and so favorably known by psychologists and by the many users of the method that no words of editorial introduction are needed to call attention to the importance of the present monograph.

The results of these labors are embodied in the Stanford Revision of the Binet-Simon Scale. A general guide for the application of this Revision has been published elsewhere. In the present monograph, however, the reader is taken "behind the scenes," is shown the precise methods by which the Revision was made, the actual data on which it was based. There is introduced also an instructive discussion of a number of very salient questions: What is the nature of intelligence? How is intelligence distributed? What sex differences exist in intelligence? What is the relation between intelligence and social status? Between intelligence and school success? Is the intelligence quotient a valid measure? How shall the validity of any single test in an intelligence scale be determined? What principles should govern the assembling of tests into a system, or scale? These questions have more than a merely technical interest: they bear in many ways upon practical problems of school instruction and administration. The monograph should do much to stimulate and to clarify thinking, both in psychological and in pedagogical circles.

G. M. W.

PREFACE

The present monograph summarizes the data on which the Stanford revision and extension of the Binet scale rests and gives an analysis of the results secured by the application of the revised scale with nearly 1000 unselected school children.

The complete guide for giving and scoring the tests and for the interpretation of results is published separately: *The Measurement of Intelligence* (Houghton Mifflin Co., 1916). This and the present monograph are in a sense companion volumes, and it is especially hoped that all who use the guide will also make themselves familiar with source material herein offered.

The responsibility of each of the various collaborators is related in Chapter I. Terman is responsible for the assembling of the source material, the arrangement of the trial series, the scoring of all the records, the elaboration of the revision from the results, the formulation of the procedure, the analysis of the data, and the preparation of this monograph for the press. In all these matters, however, invaluable help was rendered by all who collaborated in the work. Whatever merit the present revision possesses must be credited in no small degree to the loyal and painstaking work of those who assisted in the tests. Hearty thanks are also due the public-school officers, teachers, principals and superintendents for their always willing coöperation in furnishing pupils for the tests and in supplying the supplementary information called for.

LEWIS M. TERMAN.

Stanford University, June 12, 1916.

TABLE OF CONTENTS

CHAPTER I

BRIEF ACCOUNT OF THE STANFORD REVISION AND ITS HISTORY

Terman and Childs' tests of 396 children in 1910–1911 afforded data for a tentative revision and extension of the Binet 1908 scale. The most important changes introduced into the test series by that revision involved a shifting downward of most of the tests in the lower end of the scale, a shifting of several of the upper tests in the opposite direction, and the addition of the following new tests: the "ball and field" test, a graded completion test, and a graded test of vocabulary and fable interpretation.[1]

In 1911–1912 the tentative revision was applied to 310 public school children in Palo Alto, San Jose and Los Angeles. Of these, 127 were tested by Miss Helen Trost, a senior student at Stanford University, 52 by Dr. Charles Waddle, of the State Normal School, Los Angeles, and the remainder by Terman. The children were selected from each grade by arbitrary rule, according to seating when this had not been determined by scholarship, otherwise alphabetically. The schools selected were attended chiefly by children of the middle classes. The number at each age is given on p. 9.

For several reasons the results of this study did not afford satisfactory data for a further revision of the scale. The method of selecting subjects failed to give representative children at all the various ages, and too

[1] Lewis M. Terman and H. G. Childs: A Tentative Revision and Extension of the Binet-Simon Measuring Scale of Intelligence. *J. of Educ. Psych.*, Vol. 3, Feb., Mar., Apr. and May, 1912.

little attention had been given to securing uniformity of procedure. Moreover, some of the features of the Terman and Childs revision proved impracticable in actual use and showed the necessity of a more thoroughgoing revision based on more extensive data. Accordingly, a new investigation was undertaken, much more extensive than the earlier ones and more carefully planned.

The work was begun in the autumn of 1913 by Terman, Lyman and Galbreath. Miss Lyman and Miss Galbreath were at the time graduate students in the Department of Education at Stanford University. Later, Professor and Mrs. Ordahl, of the University of Nevada, and Mr. Wilford Talbert, a former graduate student of Stanford University, kindly offered to coöperate in the work. During the school year of 1913–1914 approximately 1000 public-school children were tested by Miss Lyman, Miss Galbreath, Mr. Talbert, Professor Ordahl and Mrs. Ordahl. The following year another Stanford graduate, Miss Irene Cuneo, secured further data on the lower tests by applying the revised scale to the 54 kindergarten children attending the training department of the California State Normal School at San Jose. The accompanying table shows the number of children of each age tested by each examiner.

The data secured from these 982 children made possible the revision of the scale up to the 14-year level, but left the extreme upper part of the scale still insecure. Revision of this part was fortunately made possible by the following data:

(1) Tests of 40 high-school students in San Jose and Campbell, Cal. Thirty-two of these, tested by Terman, were members of classes in "life-career study"

TABLE I

NUMBER OF CHILDREN AT EACH AGE TESTED BY THE SEVERAL EXAMINERS

Examiner	Place	4	5	6	7	8	9	10	11	12	13	14	15	16–17	Total
Lyman	Santa Barbara Los Angeles Los Gatos, Cal.	3	10	18	23	34	35	27	33	26	42	29	15	4	299
Prof. and Mrs. Ordahl	Reno, Nevada		11	27	31	31	38	29	22	22	33	33	19	8	304
Galbreath	San Jose and Mt. View, Cal.			18	23	23	28	20	10	25	11	10	5	1	176
Talbert	Oakland		15	32	15	12	12	11	14	10	12	10	7	1	151
Cuneo	Kindergarten St. Nor. School San Jose	14	18	22											54
Total at each age		17	54	117	92	100	113	87	79	83	98	82	46	14	982

and were of junior or senior grade. Their ages ranged from 17 to 20, with a median of 18+. The others, tested by Mr. Zeidler, were first and second-year students from 14 to 16 years of age. The "high school group," referred to in the statistics, included only the 32 tested by Terman.

(2) Tests of 30 business men in Palo Alto and vicinity, by Knollin and Zeidler. The men selected had had little or no formal education beyond the common school but had shown themselves ordinarily successful in the various lines of business represented in a small city.

(3) Tests of 150 "migrating unemployed" men by Mr. Knollin. These were temporary residents at a "hobo hotel" conducted at Palo Alto for transient pedestrians who were willing to work a few hours for a night's lodging and a couple of meals. The

ages ranged from 18 to 65, but were chiefly between 25 and 40. Tests of somewhat more than 100 unemployed men were also made with our trial series by Mr. Glenn Johnson, of Reed College, Portland, Oregon, who kindly loaned us his data for comparative purposes.

(4) Tests of 150 juvenile delinquents in the Whittier (Cal.) State School. These tests were made by Dr. J. H. Williams, at that time Fellow on the Buckel Foundation, Stanford University. The ages of the delinquents ranged from 10 to 21, but most were between 14 and 19.[2]

Returning now to the tests of 1000 unselected children, this part of the investigation may be described as follows:

1. We first assembled as nearly as possible all the results which had been secured for each test of the Binet scale by all the workers of all countries, including per cents passing the test at various ages, conditions under which the results were secured, method of procedure, etc. After a comparative study of these data, and in the light of results we had ourselves secured, a provisional arrangement of the tests was prepared for trial.

---

[2] As the foregoing studies of delinquents, unemployed, and business men are to be published separately by their several authors, it is not necessary to enter into them here in detail. Mr. Williams has since increased his tests of delinquent boys to nearly 500, and Miss Cuneo her tests of kindergarten children to approximately 100. More recently another Stanford University student has used the revision with 150 employees, mostly unskilled or semi-skilled. The mental ages found are given in Chapter II.

About a dozen additional studies, involving tests of nearly 1000 school children and adults, were carried out at Stanford University during the school year of 1916–1917. These studies, which will be reported in a forthcoming monograph, have sought especially to determine the validity of the Stanford Revision as a means of diagnosing a child's educability.

2. A plan was then devised for securing subjects who should be as nearly as possible representative of the several ages. The method was to select a school in a community of average social status, a school attended by all or practically all the children in the district where it was located. In order to get clear pictures of age differences, the tests were confined to children who were within two months of a birthday.[3] To avoid accidental selection, *all* the children within two months of a birthday were tested, in whatever grade enrolled (below the high school). Tests of foreign-born children, however, were eliminated in the treatment of results.

3. The children's responses were for the most part recorded verbatim. This made it possible to re-score the records according to any desired standard and thus to fit a test more perfectly to the age level assigned it.

4. The tests were made at an average rate of about fifty minutes per test. The time was rarely below 40 minutes, except with the children of four and five years. The older children and adults more often required from fifty minutes to an hour. In spite of the rather long time required for the test we are convinced that fatigue has been a negligible factor in our results. The tasks required of the child are so novel that the reserve energies are brought into play and attention is kept at high efficiency much longer than would be the case with ordinary school work.

5. As may be inferred from the time required, the testing was reasonably thorough. It is possible, however, that occasionally a success has been missed by not carrying the test high enough, or a failure missed

[3] The only exception to this was in the case of 14 five year olds, tested by Miss Cuneo.

by not going back far enough. Errors of this sort doubtless about balance in the long run, and so do not affect appreciably the distribution of mental ages. They do affect, however, the statistical treatment of the results for individual tests, and as a rule we have given the per cents passing a test only for those ages at which all the children were given the test.

6. Much attention was given to securing uniformity of procedure. A half-year was devoted to training the examiners and another half-year to the supervision of the testing.[4] In the further interests of uniformity all the records were scored by one person (Terman).

In working out a revision of the scale the guiding principle was to secure an arrangement of the tests and a standard of scoring which would cause the median mental age of the children of each age-group to coincide with the median chronological age. If the median mental age at any point in the scale was too high or too low, it was only necessary to change the location of certain of the tests, or to change the standard of scoring, until an order of arrangement and a standard of passing were found which would throw the median mental age where it belonged. We had already become convinced that no satisfactory revision of the Binet scale was possible on any theoretical considerations as to the percent of passes which an individual test ought to show in a given year in order to be considered standard for that year, although such a plan might be feasible with a scale differently founded.

[4] This statement does not apply, however, to Professor and Mrs. Ordahl, who had to rely on a 20-page guide, supplemented by a few demonstration tests and such further direction as could be given by correspondence. Mr. Talbert and Miss Cuneo had also somewhat less specific training for the tests than had the others, though both had taken a half-year course in clinical psychology.

As was to be expected, the first draft of the revision did not prove satisfactory. The scale was still too hard at some points and too easy at others. Three successive revisions were necessary, involving three separate scorings of the data and as many tabulations of the mental ages, before the desired degree of accuracy was secured.

As finally left, the scale gives a median intelligence quotient closely approximating 100 for our non-selected children of each age. The revision contains six regular tests and from one to three alternative tests in each year from 3 to 10, eight tests at year 12, six at 14, and six in each of two higher groups which are named, in order, "average adult" and "superior adult."

The tests in the two highest groups were standardized chiefly on the basis of results from 400 adults. The extension of the scale in the upper range is such that ordinarily intelligent adults, little educated, test near to what is called the "average adult" level. Adults whose intelligence is known from other sources to be superior are found to test well up to the "superior adult" level, whether they are well educated or practically unschooled. Of 30 uneducated business men, 15 tested at "average adult" (15–17), 8 at "superior adult" (17–19), 6 at "inferior adult" (14–15) and 1 at 13. Of 32 high-school students who were 16 years of age or older, 22 tested at "average adult," 5 at "superior adult," 5 at "inferior adult."

The trial arrangement of tests included, in addition to those of the Binet 1908 and 1911 series, 31 additional tests, as follows: Kuhlmann's test of discrimination of forms, two new tests of comprehension (Terman), four tests of repeating digits in reversed order (suggested by Bobertag), repeating 8 digits, test of ability to tie a bow-knot (Terman), two tests of finding similarities

(Terman), six vocabulary tests (Terman and Childs), two form-board tests (Healy and Fernald), the Healy-Fernald code test, two tests of fable interpretation (Terman and Childs), two "ball and field" tests (Terman and Childs), an "induction" test (Terman), a test of arithmetical reasoning (selected from Bonser's series), an "ingenuity" test (Terman), a test of "comprehension of physical relations" (suggested in part by Meumann), a test of observation (drawing an apple with pencil through it, suggested by an experiment of Professor Earl Barnes), and the "problem of enclosed boxes" (Terman). The test of observation proved too unsatisfactory to be included in the revision, as was true also of one of the Healy-Fernald form boards and Binet's "suggestion" and "reversed triangle" tests. Counting both regular and alternative tests, the revision contains 90 tests, as contrasted with 54 in the Binet 1911 series.

As far as possible, the original Binet tests have been retained in the form in which they were used by their author, although in a number of cases it has seemed advisable to introduce alterations either in procedure or scoring. In preparing the directions, special attention has been devoted to the difficulties encountered by inexperienced examiners in giving and scoring the tests.[5]

While it is not claimed that the revision here offered is satisfactory in every respect, the authors believe that it possesses a number of distinct advantages over other versions of the Binet scale. Among these advantages are the following:

[5] An extended guide for the giving and scoring of the individual tests and for the interpretation of test results has been published in a separate volume, *The Measurement of Intelligence* (Houghton Mifflin Co., 1916). With the latter is furnished all the necessary material for the use of the Stanford Revision.

1. Correction of the too-great ease of the original scale at its lower end and its too-great difficulty at the upper end. This correction should have the important result of tending to prevent the overlooking of borderline cases of deficiency among young children and the overestimation of deficiency among adults of somewhat inferior or borderline intelligence.[6]

2. The revision not only contains a much larger number of tests than any other series, but also brings into operation a much greater variety of mental functions. This is especially true for the upper part of the scale.

3. It is believed that the detailed directions set forth in the companion volume for giving and scoring the tests should tend materially to promote uniformity of procedure.

The following copy of the forms used in applying the tests will serve to indicate their nature.

YEAR III. (6 tests, 2 months each.)

1. Points to parts of body. (3 of 4.)
   Nose.....Eyes.....Mouth.....Hair.....
2. Names familiar objects. (3 of 5.)
   Key.....Penny.....Closed knife.....Watch.....Pencil.....
3. Pictures, enumeration or better. (At least 3 objects in one picture. "Tell me everything you can see in this picture.")
   a. Dutch Home.....
   b. Canoe.....
   c. Post Office.....
4. Gives sex. (Note form of question.).....
5. Gives last name.....
6. Repeats 6–7 syllables. (1 of 3.)
   a. "I have a little dog."
   b. "The dog runs after the cat."
   c. "In summer the sun is hot."

Al. Repeats 3 digits. (1 of 3. Order correct. Read 1 per second.)
   6–4–1..... 3–5–2..... 8–3–7.....

YEAR IV. (6 tests, 2 months each.)

1. Compares lines. (3 of 3, or 5 of 6.) 1.....2.....3.....
2. Discrimination of forms. (Kuhlmann. 7 of 10.)
   Circle..... Square..... Triangle.....Other errors.....

[6] For a fuller discussion, see L. M. Terman, Some problems related to the detection of border-line cases of mental deficiency, *J. of Psycho-Asthenics*, Sept. and Dec., 1915.

3. Counts 4 pennies. (No error.).....
4. Copies square. (Pencil. 1 of 3.) 1.....2.....3.....
5. Comprehension, 1st degree. (2 of 3.) "What must you do:
   a. "When you are sleepy?.....
   b. "When you are cold?.....
   c. "When you are hungry?".....
6. Repeats 4 digits. (1 of 3. Order correct. Read 1 per second.)
   4–7–3–9..... 2–8–5–4..... 7–2–6–1.....

Al. Repeats 12–13 syllables. (1 of 3 absolutely correct, or 2 with 1 error each.)
   a. "The boy's name is John. He is a very good boy."
   b. "When the train passes you will hear the whistle blow."
   c. "We are going to have a good time in the country."

YEAR V. (6 tests, 2 months each.)

1. Comparison of weights. (2 of 3.)
   3–15..... 15–3..... 3–15.....
2. Colors. (No error.)
   Red..... Yellow..... Blue..... Green.....
3. Aesthetic comparison. (No error.)
   Upper pair..... Middle..... Lower.....
4. Definitions, use or better. (4 of 6.)
   Chair..... Doll.....
   Horse..... Pencil.....
   Fork..... Table.....
5. Patience, or divided rectangle. (2 of 3 trials. 1 minute each.)
   1..... Time.....
   2..... Time.....
   3..... Time.....
6. Three commissions. (No error. Order correct.)
   Puts key on chair..... Brings box..... Shuts door.....

Al. Age.....

YEAR VI. (6 tests, 2 months each.)

1. Right and left. (3 of 3, or 5 of 6.)
   R. hand..... L. ear..... R. eye.....
2. Mutilated pictures. (3 of 4.)
   Eye..... Mouth..... Nose..... Arms.....
3. Counts 13 pennies. (1 of 2 trials, without error.).....
4. Comprehension, 2d degree. (2 of 3.) "What's the thing to do:
   a. "If it is raining when you start to school?.....
   b. "If you find that your house is on fire?.....
   c. "If you are going some place and miss your car?".....
5. Coins. (3 of 4. Present in order given below.)
   Nickel..... Penny..... Quarter..... Dime.....
6. Repeats 16–18 syllables. (1 of 3 absolutely correct, or 2 with 1 error each.)
   a. "We are having a fine time. We found a little mouse in the trap."
   b. "Walter had a fine time on his vacation. He went fishing every day."
   c. "We will go out for a long walk. Please give me my pretty straw hat."

Al. Morning or afternoon. (Note form of question.).....

YEAR VII. (6 tests, 2 months each.)

1. Fingers. (No error.) R..... L..... Both.....
2. Pictures, description or better. (Over half of performance description. "Tell me what this picture is about?" "What is this a picture of?")
   a. Dutch Home.....
   b. Canoe.....
   c. Post Office.....
3. Repeats 5 digits. (1 of 3. Order correct. Read 1 per second.)
   3–1–7–5–9..... 4–2–8–3–5..... 9–8–1–7–6.....
4. Ties bow knot. (Model shown. 1 minute. "Single" bow half credit.)
   Time..... Method.....
5. Gives differences. (2 of 3.)
   a. Fly and butterfly.....
   b. Stone and egg.....
   c. Wood and glass.....
6. Copies diamond. (Pen. 2 of 3.) *a*.....*b*.....*c*.....

Al. 1. Names days of week. (Order correct. 2 of 3 checks correct.)
Mon., Tues., Wed., Thurs., Fri., Sat., Sun.

Al. 2. Repeats 3 digits backwards. (1 of 3. Read 1 per second.)
2–8–3..... 4–2–7..... 9–5–8.....

YEAR VIII. (6 tests, 2 months each.)

1. Ball and field. (Inferior plan or better.).....
2. Counts 20–0. (40 seconds. 1 error allowed.) Time....Errors....
3. Comprehension, 3rd degree. (2 of 3.) "What's the thing for you to do:
   a. "When you have broken something which belongs to someone else?.....
   b. "When you are on your way to school and notice that you are in danger of being tardy?.....
   c. "If a playmate hits you without meaning to do it?".....
4. Gives similarities, two things. (2 of 4. "In what way are wood and coal alike?" etc. Any real likeness is plus.)
   a. Wood and coal.....
   b. Apple and peach.....
   c. Iron and silver.....
   d. Ship and automobile.....
5. Definitions superior to use. (2 of 4. "Thing" as genus counts plus.)
   a. Balloon.....
   b. Tiger.....
   c. Football.....
   d. Soldier.....
6. Vocabulary, 20 words. Score..... Total Vocab.....

Al. 1. Six coins. (No error. Give in order indicated.)
.05..... .01..... .25..... .10..... 1.00..... .50.....

Al. 2. Dictation. ("See the little boy." Easily legible. Pen, 1 minute.)
Time..... Score by Ayres scale.....

YEAR IX. (6 tests, 2 months each.)

1. Date. (Allow error of 3 days in *c*, no error in *a*, *b*, or *d*.)
   a. Day of week.... *b*. month.... *c*. day of m..... *d*. year.....
2. Weights. (3, 6, 9, 12, 15. Procedure not illustrated. 2 of 3 correct.)
   a. ..... Method.....
   b. ..... Method.....
   c. ..... Method.....
3. Makes change. (2 of 3. No coins, paper, or pencil.)
   10–4..... 15–12..... 25–4.....
4. Repeats 4 digits backwards. (1 of 3. Read 1 per second.)
   6–5–2–8..... 4–9–3–7..... 8–6–2–9.....
5. Three words. (2 of 3. Oral. 1 sentence or not over 2 coördinate clauses.)
   a. Boy, river, ball.....
   b. Work, money, men.....
   c. Desert, rivers, lakes.....
6. Rhymes. (3 rhymes for each word. 1 minute for each part. Illustrate with hat, rat, cat.)
   a. Day..... Time.....
   b. Mill..... Time.....
   c. Spring..... Time.....

Al. 1. Months. (15 seconds and 1 error in naming. 2 checks of 3 correct.)
Jan., Feb., Mch., Apr., May, June, July, Aug., Sept., Oct., Nov., Dec.

Al. 2. Stamps, gives total value. (2d trial if individual values are known.)

YEAR X. (6 tests, 2 months each.)

1. Vocabulary, 30 words. Score..... Total Vocab.....
2. Absurdities. (4 of 5. Warn. Spontaneous correction allowed.)
   a. "A man said: 'I know a road from my house to the city which is down hill all the way to the city and down hill all the way back home.'"
   b. "An engineer said that the more cars he had on his train the faster he could go."
   c. "Yesterday the police found the body of a girl cut into 18 pieces. They believe that she killed herself."
   d. "There was a railroad accident yesterday, but it was not very serious. Only 48 people were killed."
   e. "A bicycle rider, being thrown from his bicycle in an accident, struck his head against a stone and was instantly killed. They picked him up and carried him to the hospital, and they do not think he will get well again."
3. Designs. (1 correct, 1 half correct. Expose 10 seconds.) *a*...*b*...
4. Reading and report. (8 memories. 35 seconds and 2 mistakes in reading.)
   Memories..... Time for reading..... Mistakes.....

New York. | September 5th. | —A fire | last night | burned | three houses | near the center | of the city. | It took some time | to put it out. |

The loss | was fifty thousand dollars, | and seventeen families | lost their homes. | In saving | a girl | who was asleep | in bed, | a fireman | was burned | on the hands.

5. Comprehension, 4th degree. (2 of 3. Question may be repeated.)
   a. "What ought you to say when someone asks your opinion about a person you don't know very well?".....
   b. "What ought you to do before undertaking (beginning) something very important?".....
   c. "Why should we judge a person more by his actions than by his words?".....
6. 60 words. (Score half-minutes separately. Illustrate with clouds, dog, chair, happy.) 1.....2.....3.....4.....5.....6.....
   Method.....

Al. 1. Repeats 6 digits. (1 of 2. Order correct. Read 1 per second.)
3–7–4–8–5–9..... 5–2–1–7–4–6.....

Al. 2. Repeats 20–22 syllables. (1 of 3 correct, or 2 with 1 error each.)
   a. "The apple tree makes a cool pleasant shade on the ground where the children are playing."
   b. "It is nearly half-past one o'clock; the house is very quiet and the cat has gone to sleep."
   c. "In summer the days are very warm and fine; in winter it snows and I am cold."

Al. 3. Form board. (Healy-Fernald Puzzle A. 3 times in 5 minutes.)
Time: *a*.....*b*.....*c*..... Method.....

YEAR XII. (8 tests, 3 months each.)

1. Vocabulary, 40 words. Score..... Total Vocab.....
2. Abstract words. (3 of 5.)
   a. Pity.....
   b. Revenge.....
   c. Charity.....
   d. Envy.....
   e. Justice.....
3. Ball and field. (Superior plan.).....
4. Dissected sentences. (2 of 3. 1 minute each.)
   a. ..... Time.....
   b. ..... Time.....
   c. ..... Time.....
5. Fables. (Score 4, *i. e.*, two correct or the equivalent in half credits.)
   a. Hercules and wagoner.....
   b. Maid and eggs.....
   c. Fox and crow.....
   d. Farmer and stork.....
   e. Miller, son and donkey.....
6. Repeats 5 digits backwards. (1 of 3. Read 1 per second.)
   3–1–8–7–9..... 6–9–4–8–2..... 5–2–9–6–1.....
7. Pictures, interpretation. (3 of 4. "Explain this picture.")
   a. Dutch Home.....
   b. Canoe.....

c. Post Office.....
d. Colonial Home.....
8. Gives similarities, three things. (3 of 5. "In what way are —, —, —, alike?" Grade fairly closely.)
a. Snake, cow, sparrow.....
b. Book, teacher, newspaper.....
c. Wool, cotton, leather.....
d. Knife-blade, penny, piece of wire.....
e. Rose, potato, tree.....

YEAR XIV. (6 tests, 4 months each.)

1. Vocabulary, 50 words. Score..... Total Vocab.....
2. Induction test. (Gets rule by 6th folding. Unfold after each cutting.)
1..... 2..... 3..... 4..... 5..... 6.....
3. President and king. (Power.... accession.... tenure.... 2 of 3.)
a. .....
b. .....
c. .....
4. Problems of fact.. (2 of 3. Query on *a* and *b*.)
a. "A man who was walking in the woods near a city stopped suddenly, very much frightened, and then ran to the nearest policeman, saying that he had just seen hanging from the limb of a tree a——a what?".....
.....
b. "My neighbor has been having queer visitors. First a doctor came to his house, then a lawyer, then a minister (preacher or priest). What do you think happened there?"
.....
c. "An Indian who had come to town for the first time in his life saw a white man riding along the street. As the white man rode by the Indian said—'The white man is lazy; *he walks sitting down.*' What was the white man riding on that caused the Indian to say 'he walks sitting down'?"
.....
5. Arithmetical reasoning. (1 minute each. 2 of 3.)
a. If a man's salary is $20 a week and he spends $14 a week, how long will it take him to save $300?
.....
b. If 2 pencils cost 5 cents, how many pencils can you buy for 50 cents?.....
c. At 15 cents a yard, how much will 7 feet of cloth cost?.....
6. Clock. (2 of 3. Error must not exceed 3 or 4 minutes.)
6:22..... Time required.....
8:10..... Time required.....
2:46..... Time required.....
Al. Repeats 7 digits. (1 of 2. Order correct. Read 1 per second.)
2–1–8–3–4–3–9..... 9–7–2–8–4–7–5.....

YEAR XVI, AVERAGE ADULT. (6 tests, 5 months each.)

1. Vocabulary, 65 words. Score..... Total Vocab.....
2. Interpretation of fables. (Score 8.) (First explain what a fable is, and after reading each say, "What lesson does that teach us?")
   a. Hercules and wagoner.....
   b. Maid and eggs.....
   c. Fox and crow.....
   d. Farmer and stork.....
   e. Miller, son and donkey.....
3. Difference between abstract words. (3 real contrasts out of 4.)
   a. Laziness and idleness.....
   b. Evolution and revolution.....
   c. Poverty and misery.....
   d. Character and reputation.....
4. Problem of the enclosed boxes. (3 of 4.) One large box containing:
   a. 2 smaller, 1 inside of each.....
   b. 2 smaller, 2 inside of each.....
   c. 3 smaller, 3 inside of each.....
   d. 4 smaller, 4 inside of each.....
5. Repeats 6 digits backwards. (1 of 3.)
   4–7–1–9–5–2..... 5–8–3–2–9–4..... 7–5–2–6–3–8.....
6. Code, writes "Come quickly." (2 errors. Omission of dot counts half error. Illustrate with "war," "trench," and "spy.")
   Errors.....C-O-M-E Q-U-I-C-K-L-Y..... Time.....
   Method.....

Al. 1. Repeats 28 syllables. (1 of 2 absolutely correct.)
   a. Walter likes very much to go on visits to his grandmother, because she always tells him many funny stories.
   b. Yesterday I saw a pretty little dog in the street. It had curly brown hair, short legs, and a long tail.

Al. 2. Comprehension of physical relations. (2 of 3.)
   a. Path of cannon ball.....
   .....
   b. Weight of fish in water.....
   .....
   c. Hitting distant mark.....
   .....

XVIII, SUPERIOR ADULT. (6 tests, 6 months each.)

1. Vocabulary, 75 words. Score..... Total Vocab.....
2. Binet's paper cutting test. Draws folds and locates holes. (If given, must come before XIV_2.)
3. Repeats 8 digits. (1 of 3. Order correct. Read 1 per second.)
   7–2–5–3–4–8–9–6.... 4–9–8–5–3–7–6–2.... 8–3–7–9–5–4–8–2....
4. Repeats thought of passage heard. (1 of 2. E. reads each in about ½ min.) "I am going to read you a little selection. Listen carefully, and when I am through I will ask you to tell as much of it as you can remember. Ready—"

a. "Tests such as we are now making are of value both for the advancement of science and for the information of the person who is tested. It is important for science to learn how people differ and on what factors these differences depend. If we can separate the influence of heredity from the influence of environment we may be able to apply our knowledge so as to guide human development. We may thus in some cases correct defects and develop abilities which we might otherwise neglect."

.....
.....
.....
.....
.....
.....
.....

b. "Many opinions have been given on the value of life. Some call it good, other call it bad. It would be nearer correct to say that it is mediocre, for on the one hand our happiness is never as great as we should like, and on the other hand our misfortunes are never as great as our enemies would wish for us. It is this mediocrity of life which prevents it from being radically unjust."

.....
.....
.....
.....
.....
.....
.....

5. Repeats 7 digits backwards. (1 of 3.)
   4–1–6–2–5–9–3..... 3–8–2–6–4–7–5..... 9–4–5–2–8–3–7.....
6. Ingenuity test. (2 of 3. 5 minutes each. If S fails on 1st, E explains that one.)
   a. "A mother sent her boy to the river to get seven pints of water. She gave him a 3-pint vessel and a 5-pint vessel. Show me how the boy can measure out exactly 7 pints without guessing at the amount. Begin by filling the 5-pint vessel."

   .....
   .....

   b. Same, except 5 and 7 given to get 8. ("Begin with 5.")

   .....
   .....

   c. Same, except 4 and 9 given to get 7. ("Begin with 4.")

   .....
   .....

THE VOCABULARY TEST

Time required.....

1. orange.....
2. bonfire.....
3. roar.....
4. gown.....
5. tap.....
6. scorch.....
7. puddle.....
8. envelope.....
9. straw.....
10. rule.....
11. haste.....
12. afloat.....
13. eye-lash.....
14. copper.....
15. health.....
16. curse.....
17. guitar.....
18. mellow.....
19. pork.....
20. impolite.....
21. plumbing.....
22. outward.....
23. lecture.....
24. dungeon.....
25. southern.....
26. noticeable.....
27. muzzle.....
28. quake.....
29. civil.....
30. treasury.....
31. reception.....
32. ramble.....
33. skill.....
34. misuse.....
35. insure.....
36. stave.....
37. regard.....
38. nerve.....
39. crunch.....
40. juggler.....
41. majesty.....
42. brunette.....
43. snip.....
44. apish.....
45. sportive.....
46. hysterics.....
47. Mars.....
48. repose.....
49. shrewd.....
50. forfeit.....

Score.....

51. peculiarity.....
52. coinage.....
53. mosaic.....
54. bewail.....
55. disproportionate.....
56. dilapidated.....
57. charter.....
58. conscientious.....
59. avarice.....
60. artless.....
61. priceless.....
62. swaddle.....
63. tolerate.....
64. gelatinous.....
65. depredation.....
66. promontory.....
67. frustrate.....
68. milksop.....
69. philanthropy.....
70. irony.....
71. lotus.....
72. drabble.....
73. harpy.....
74. embody.....
75. infuse.....
76. flaunt.....
77. declivity.....
78. fen.....
79. ochre.....
80. exaltation.....
81. incrustation.....
82. laity.....
83. selectman.....
84. sapient.....
85. retroactive.....
86. achromatic.....
87. ambergris.....
88. casuistry.....
89. paleology.....
90. perfunctory.....
91. precipitancy.....
92. theosophy.....
93. piscatorial.....
94. sudorific.....
95. parterre.....
96. homunculus.....
97. cameo.....
98. shagreen.....
99. limpet.....
100. complot.....

NOTE: To get the entire vocabulary, multiply the number of correct definitions by 180.

The Fable Test

"Fables, you know, are little stories which teach us a lesson. Now I am going to read a fable to you. Listen carefully and when I am through I will ask you to tell what lesson the fable teaches us."

After reading each fable say, "What lesson does that teach us?" Ask also if fable has been heard before.

A. Hercules and the Wagoner

A man was driving along a country road, when the wheels suddenly sank in a deep rut. The man did nothing but look at the wagon and call loudly to Hercules to come and help him. Hercules came up, looked at the man, and said: "Put your shoulder to the wheel, my man, and whip up your oxen." Then he went away and left the driver.
Lesson.....

B. The Milkmaid and Her Plans

A milkmaid was carrying her pail of milk on her head, and was thinking to herself thus: "The money for this milk will buy 4 hens; the hens will lay at least 100 eggs; the eggs will produce at least 75 chicks; and with the money which the chicks will bring I can buy a new dress to wear instead of the ragged one I have on." At this moment she looked down at herself, trying to think how she would look in her new dress; but as she did so the pail of milk slipped from her head and dashed upon the ground. Thus all her imaginary schemes perished in a moment.
Lesson.....

C. The Fox and the Crow

A crow, having stolen a bit of meat, perched in a tree and held it in her beak. A fox, seeing her, wished to secure the meat, and spoke to the crow thus: "How handsome you are! and I have heard that the beauty of your voice is equal to that of your form and feathers. Will you not sing for me, so that I may judge whether this is true?" The crow was so pleased that she opened her mouth to sing and dropped the meat, which the fox immediately ate.
Lesson.....

D. The Farmer and the Stork

A farmer set some traps to catch cranes which had been eating his seed. With them he caught a stork. The stork, which had not really been stealing, begged the farmer to spare his life, saying that he was a bird of excellent character, that he was not at all like the cranes, and that the farmer should have pity on him. But the farmer said: "I have caught you with these robbers, the cranes, and you have got to die with them."
Lesson.....

E. The Miller, His Son, and the Donkey

A miller and his son were driving their donkey to a neighboring town to sell him. They had not gone far when a child saw them and cried out: "What fools those fellows are to be trudging along on foot when one of them might be riding." The old man, hearing this, made his son get on the donkey, while he himself walked. Soon they came upon some men. "Look," said one of them, "see that lazy boy riding while his old father has to walk." On hearing this the miller made his son get off, and he climbed upon the donkey himself. Farther on they met a company of women, who shouted out: "Why, you lazy old fellow, to ride along so comfortably while your poor boy there can hardly keep pace by the side of you!" And so the good-natured miller took his boy up behind him and both of them rode. As they came to the town a citizen said to them, "Why, you cruel fellows! you two are better able to carry the poor donkey than he is to carry you." "Very well," said the miller, "we will try." So both of them jumped to the ground, got some ropes, tied the donkey's legs to a pole and tried to carry him. But as they crossed the bridge the donkey became frightened, kicked loose and fell into the stream.

Lesson.....

CHAPTER II

THE DISTRIBUTION OF INTELLIGENCE

The question as to the manner in which intelligence is distributed relates itself at once to fundamental issues in biological theory and suggests social and educational problems of great importance. Perhaps the most vital question which can be asked by any nation of any age is the following: "How high is the average level of mental endowment among our people, and how frequent are the various grades of ability above and below the average?"

With the development of standardized intelligence tests we are approaching, for the first time, a possible answer to this question. The future of such tests is guaranteed by the importance of the problems which they undertake to answer. This would still be true even if it could be shown that all the mental tests which have yet been devised or suggested are of little worth.

Difficulties in Finding the True Distribution of Intelligence

In view of the large number of investigations made with the Binet-Simon tests in many countries, the light which these have thrown upon the distribution of intelligence is less than might have been expected. The reasons for this are various. In the first place, the number of children tested by any one investigator, and particularly the number at any one age, has usually fallen short of that required for far-reaching conclusions. If we could mass the results of different investigators the problem would be made much easier;

but owing to the lack of uniformity in the methods by which the data have been secured, this is usually a dangerous procedure. Because of the small numbers we can seldom be sure that the children tested were representative. Educational advantages, social status, racial differences, and other possible selective influences must be taken into account. To get representative children of a given age is especially difficult, and it is by no means easy even when questions of education, social status and race have been eliminated. Studies of the progress of school children through the grades have shown that children of any given age are scattered over an astonishing range of grades. It has been a common mistake to select certain school grades for the testing and to suppose that the results secured could be used as norms for the ages found in those grades.

One factor or another has entered to impair the value of almost every experiment with the scale. Kuhlmann, for example, in his tests of 1000 children, avoided selection by examining all the children enrolled in the public schools of a small middle-class city; but his examiners were untrained. Fewer tests by trained examiners would have made his experiment of greater value in several respects. Binet's 1908 scale was based on tests of only 200 children, 15 to 25 of each age, and these were situated in one of the poorest quarters of Paris. What further selection of subjects took place in this experiment we are not informed. Bobertag's subjects were in the main pupils attending the *Volksschule*, and these are known to have a lower average level of mental endowment than pupils attending the higher schools. Goddard's numbers were fairly large, but the tests were made by persons of limited training and at a rate (as high as thirty per day for one tester)

which rendered thoroughness impossible. Some of the tests of Terman and Childs were made by only partly-trained examiners, and here, again, too little attention was paid to social class and age selection.

In like manner it would be easy to point out serious shortcomings in every study which has been made with the Binet scale, including those of Jeronutti at Rome, Trèves and Saffiotti at Milan, Dr. Anna Schubert at St. Petersburg, Mrs. Wolkowitsch at Moscow, Bloch and Preiss at Kattowitz, Miss Johnston at Sheffield, Winch in London, Rogers and McIntyre at Aberdeen, Decroly at Brussels, Levistre and Morlè in Paris, Miss Dougherty in Kansas City, Dr. Morse and Miss Strong in South Carolina, Dr. Rowe in Michigan, the Weintrobs in New York City, Dr. Schmidt in Chicago, etc.

But even had the procedure and the method of selecting cases not been at fault in these studies, the results would still have been misleading as regards the distribution of intelligence, owing to the acknowledged imperfection of the scale used. The arrangement of tests in the earlier years magnified decidedly the amount and range of superior ability, and to a corresponding degree covered up the presence of retardation. At the upper end of the scale the tests were so few and so unsatisfactory that individual differences above the level of eleven years were hardly brought out at all. Only at the middle point was the scale reasonably accurate.

As regards the distribution of intelligence we believe that the Stanford 1914–1915 data have more than ordinary significance, and for reasons which it may be well to enumerate:

(1) The children were as nearly representative of the different ages as it is possible to get. The method

was to select a school attended by all the children of school age in the community, and to test the children of the various ages in whatever grades they might happen to be. This obviates accidental age selection at least for the years 7 to 13, inclusive. It is possible, however, that six-year-old and fourteen-year-old school children are not quite representative of children of those ages, since mentally retarded six-year-olds are likely to enter school a year late and since a considerable number of fourteen-year-olds have been promoted to the high school.

(2) The children tested were all within two months of a birthday. Our curve of distribution for nine-year intelligence, for example, really represents the distribution of intelligence for nine-year olds, and not that of children ranging all the way from eight and a half to nine and a half years.

(3) The schools selected for the tests were such as almost any one would classify as middle-class. Few children attending them were either from very wealthy or very poor homes. The only exception to this rule was in the case of Reno, where all the children within two months of a birthday were tested throughout the city. The large majority of these, however, were from homes of average wealth and culture. Only 2 per cent., in fact, were classified by the teachers as of "very inferior" social status, and only 1.6 per cent. as of "very superior" social status.

(4) Care was taken to avoid racial differences and the difficulties due to lack of familiarity with the language. None of the children was foreign-born and only a few were of other than western European descent. The names were chiefly English, Irish, Scotch and German, with a few Swedish, French, Spanish, Italian, and Portuguese. Tests of Spanish,

Italian and Portuguese children were eliminated from our study of distribution, for the reason that in western cities children of these nationalities are likely to belong to unfavorably selected classes. We are justified in believing, therefore, that the distribution of intelligence among our subjects is less influenced by extraneous factors than has been the case in other studies of this kind.

(5) The numbers tested were relatively large, namely 80 to 120 at each age from 6 to 14 years, with somewhat fewer at 5 and 15. Moreover, the use of only such children as were within two months of a birthday greatly enhances the value of the numbers employed. In each school this near-birthday group composed approximately one-third of all the children in attendance, and by the laws of chance we can be sure that the results are approximately as they would have been had the entire school enrollment been tested, that is, 3000 instead of 1000. This plan has the further advantage that three times as many schools were sampled as would otherwise have been the case; and if it happened, in spite of the care exercised, that some of the schools were below average in social status, this would probably be counterbalanced by other schools somewhat above average.

(6) The scale by which the mental ages were computed is certainly much more accurate than Binet's or the earlier revisions. As already stated, the tests were worked over until the average mental age secured at each level approximately coincided with the physical age. In this way the scale was made of nearly equal accuracy at every point.

(7) Correcting the rather large error of the scale at the upper and lower ranges gives another advantage of extreme importance, for it enables us to combine

the intelligence quotients of the children of all ages into a single surface of distribution. As long as intelligence was reckoned in terms of years and months of retardation or acceleration it was of course not permissible to combine the distributions for different ages. The range of mental ages is approximately twice as great at 10 years as at 5 years, so that one year of retardation or acceleration at 5 is equivalent to two years of retardation or acceleration at 10. Accordingly, to combine the results for children of several different ages so as to show what percent of the entire number are retarded or accelerated one year, two years, three years, etc., is an absurd procedure. The range of intelligence quotients, however, as measured by the revised scale, is not far from constant from five to fourteen years, and these may therefore be combined into a single surface of distribution. The curve thus obtained differs from those of the individual years only in being somewhat more regular.

(8) Finally, though by no means least in importance, the tests from which the following curves of distribution were derived were made with more than ordinary care. The examiners were trained, the procedure was kept as uniform as possible and the scoring was all done by one person. Wherever evidence was found of mistaken procedure, the examiner was questioned, and if the records of that examiner for that test were not comparable with the others, they were thrown out.

After making the necessary eliminations because of incomplete testing, foreign parentage, etc., there remained 981 children, distributed as shown herewith.

TABLE 2

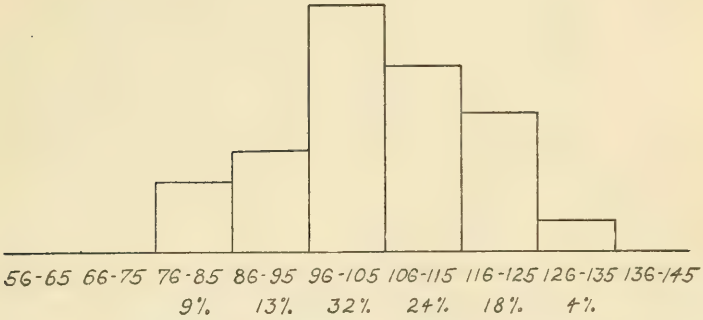
Age	4	5	6	7	8	9	10	11	12	13	14	15	16
Cases	16	54	117	92	100	113	87	97	83	98	82	46	14

The Distribution of Intelligence for the Ages Separately

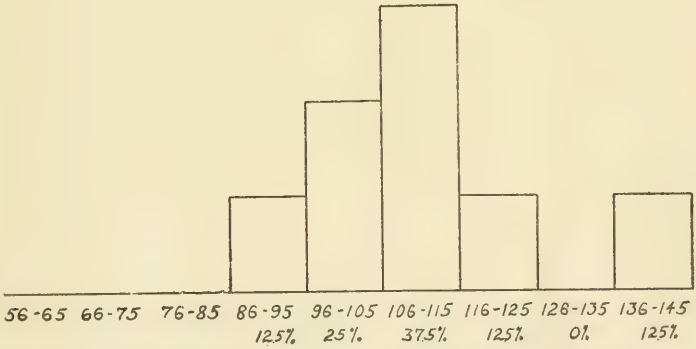
The intelligence quotients were calculated for all the children and those for a given year were then thrown into groups so that each group represented the cases include in a range of ten points intelligence quotient. The middle group includes the intelligence quotients from 96 to 105; the ascending groups include in order those from 106–115, 116–125, 126–135, and 136–145. The corresponding descending groups are 86–95, 76–85, 66–75, and 56–65. Only one case tested below 56, a girl of 8 years who had an intelligence quotient of 51. None tested above 145. Graphs 1 to 13 show the results of this grouping for each age separately from 4 to 16.

It is evident at a glance at these graphs that the distribution of intelligence quotients is fairly symmetrical at each age from 5 to 14. At 14 the number of exceptionally superior children decreases, as we should expect, since a few of the brightest have by that age finished the eighth grade. At 15 the intelligence quotients range on either side of 90 as a median, and at 16 years on either side of 80 as a median. This, again, is what we should expect, because we know from other facts that a majority of school retardates are below average in intelligence. The 17 4-year-olds averaged high, with a median intelligence quotient of 103. It is reasonably certain, however, that children of this age attending school are usually somewhat beyond the average in intelligence.

It will be noted that at every year from 5 to 14 the middle group is the largest and that the groups on either side decrease in size somewhat regularly with increasing distance from the median. In the lower years, however, the 106–115 and 116–125 groups are larger than the 86–95 and 76–85 groups respectively

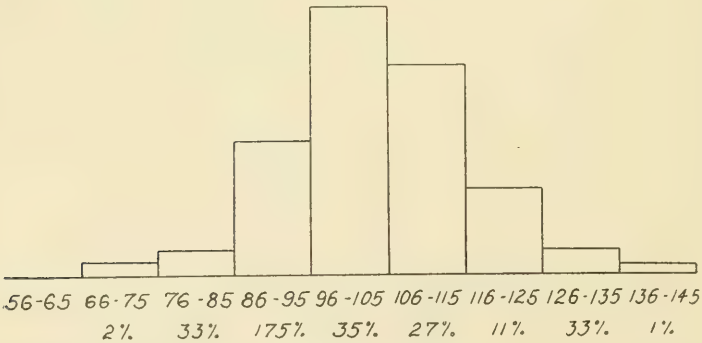


GRAPH 1. Distribution of intelligence quotients of 16 kindergarten children, age 4 years. (These pupils are a selected group and test high.)

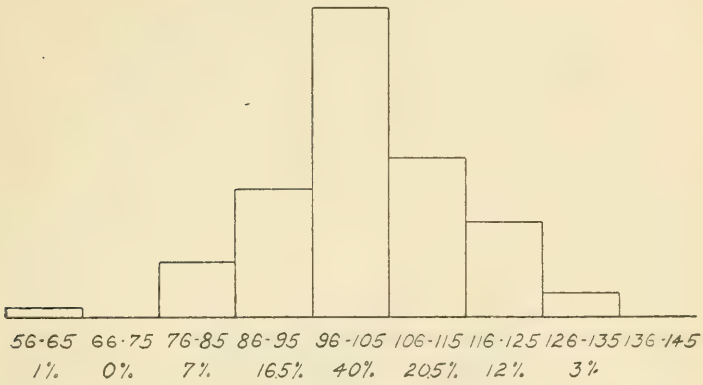


GRAPH 2. Distribution of intelligence quotients of 54 kindergarten children, age 5 years. (Median at 102.)

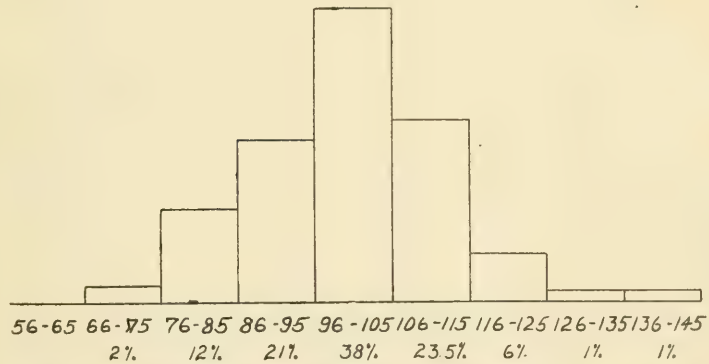
GRAPH 3. Distribution of intelligence quotients of 117 unselected 6-year-olds. (Median at 103.)



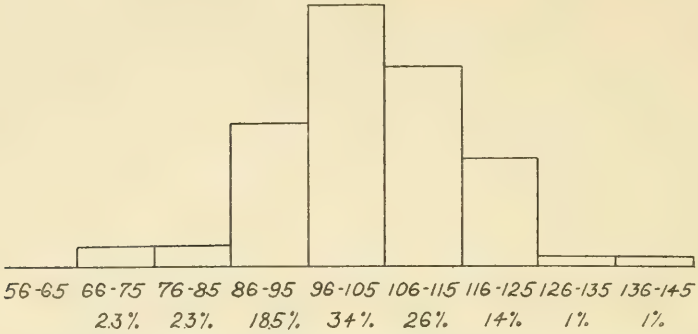
GRAPH 4. Distribution of intelligence quotients of 93 unselected 7-year-olds. (Median at 102.)



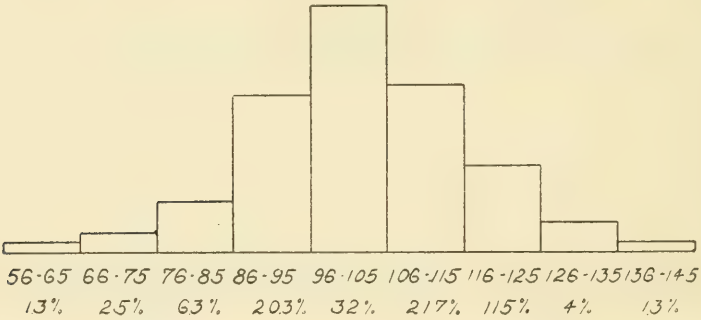
GRAPH 5. Distribution of intelligence quotients of 98 unselected 8-year-olds. (Median at 101+.)



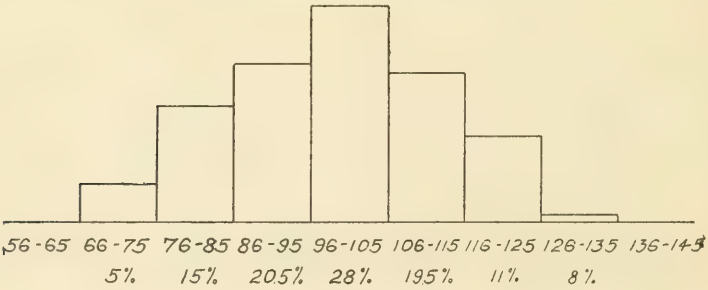
GRAPH 6. Distribution of intelligence quotients of 113 unselected 9-year-olds. (Median at 100.5.)



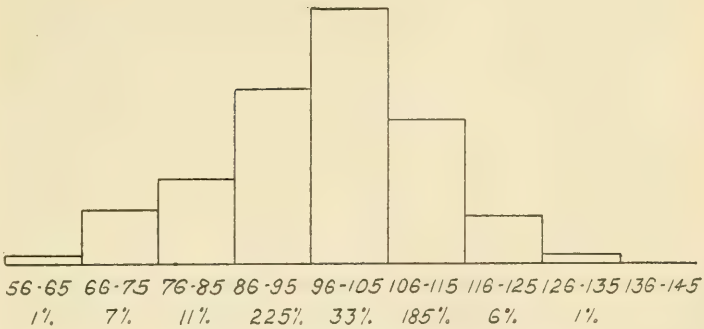
GRAPH 7. Distribution of intelligence quotients of 87 unselected 10-year-olds. (Median at 103.)



GRAPH 8. Distribution of intelligence quotients of 79 unselected 11-year-olds. (Median at 98.)

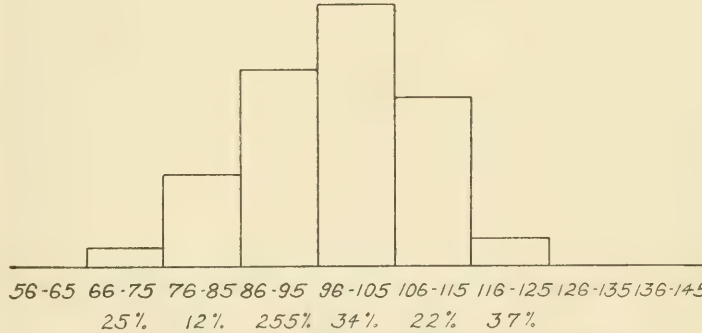


GRAPH 9. Distribution of intelligence quotients of 83 unselected 12-year-olds. (Median at 98.)



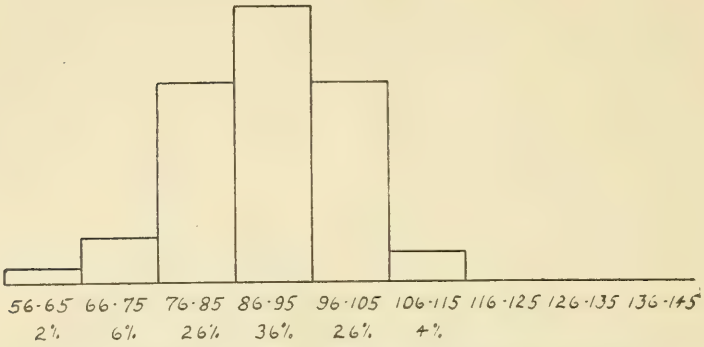
GRAPH 10. Distribution of intelligence quotients of 98 unselected 13-year-olds. (Median at 96.5.)

A few of the brightest 13-year-olds had been eliminated from this group by promotion to the high school and were not tested.



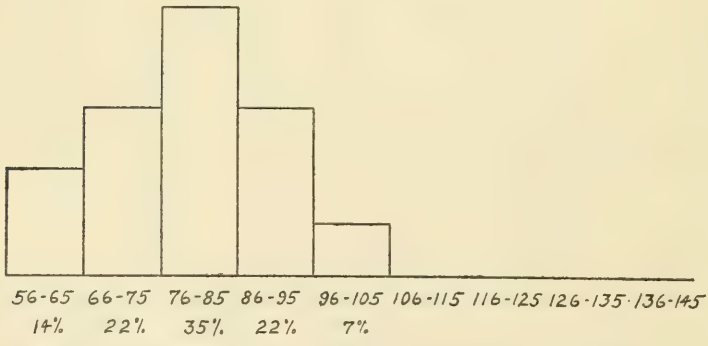
GRAPH 11. Distribution of intelligence quotients of 82 children 14 years of age. (Median at 97+.)

Many of the brightest 14-year-olds had been promoted to the high school and do not appear in the above graph.



GRAPH 12. Distribution of intelligence quotients of 47 children, age 15 years (over-age for grade). (Median 90.)

Most of the average and superior 15-year-olds had been promoted to the high school and do not appear in the above graph.



GRAPH 13. Distribution of intelligence quotients of 14 children, age 16 years (over-age for grade). (Median 80.)

All of the average and superior 16-year-olds had been promoted to the high school and do not appear in the above graph.

on the other side. The most noticeable lack of symmetry occurs at ages 5, 6, and 14. This is due in part to a certain amount of unavoidable selection. The five-year-olds were enrolled in kindergartens, and since school attendance at this age is not compulsory, we can not be sure that kindergarten children represent the median intelligence for five-year-olds. The same is true of six-year-olds, though to a less extent. In both cases the distribution of intelligence quotients suggests that at these ages inferior children are somewhat less likely to be found in school than those of superior endowment. The reverse is the case at 14, since a few of the brighter children of that age have completed the eighth grade and have either dropped out or passed on to the high school. It is possible that a few children under 14 have managed to evade the compulsory attendance laws and are not in school, but it is certain that in the cities and towns where our testing was done the amount of such evasion was practically negligible.

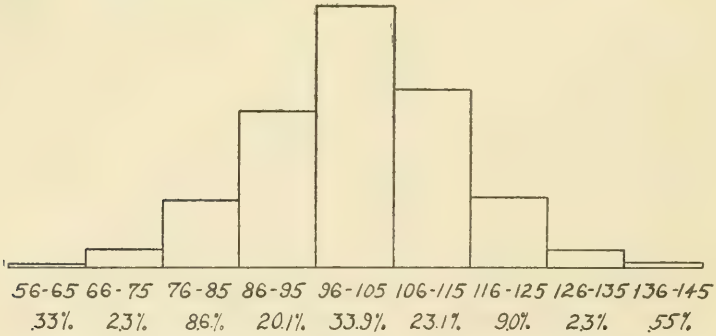
On the whole, it is evident that the distribution of intelligence quotients is fairly regular and uniform in the various years. This is further shown by the fact that the range including the middle 50 percent of the intelligence quotients does not vary greatly between 5 and 14 years. Combining each two successive years in order to overcome the chance effects of the limited number of children of any one age, we have the accompanying table in which, it is of interest to note, the distribution of intelligence quotients for the different age-levels furnishes no support to the very generally accepted belief that variability materially increases at adolescence. As far as 14, at least, there is no evidence that this occurs.

TABLE 3

Ages	Limits of quotient including middle 50%	Range of quotients including middle 50%
5 and 6 combined	97 to 111	15
7 and 8 combined	95+ to 111	16+
9 and 10 combined	94 to 108	15
11 and 12 combined	92 to 108	17
13 and 14 combined	90 to 105	16
All ages combined	92+ to 108	16+

The Distribution of Intelligence for the Ages Combined

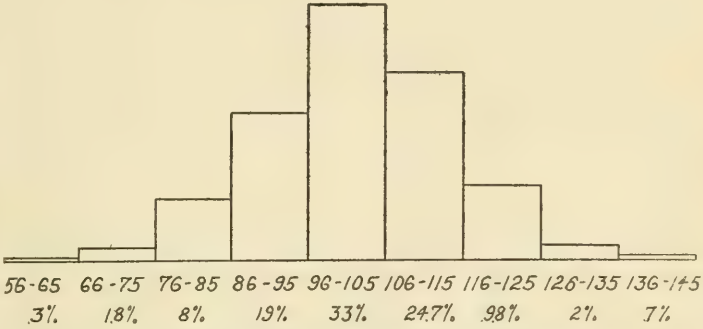
As already explained, the corrected scale has the advantage of enabling us to combine the intelligence quotients of the children of different ages into a single surface of distribution, something that we could not do when the scale was too easy at some levels and too hard at others. By combining ages 5 to 14, we have the distribution shown in Graph 14. Ages 4, 15 and 16 have been omitted from this combined distribution because of the selection which has taken place at these years.[1]



GRAPH 14. Distribution of intelligence quotients of 905 unselected children, ages 5–14 years. (Median at 99.)

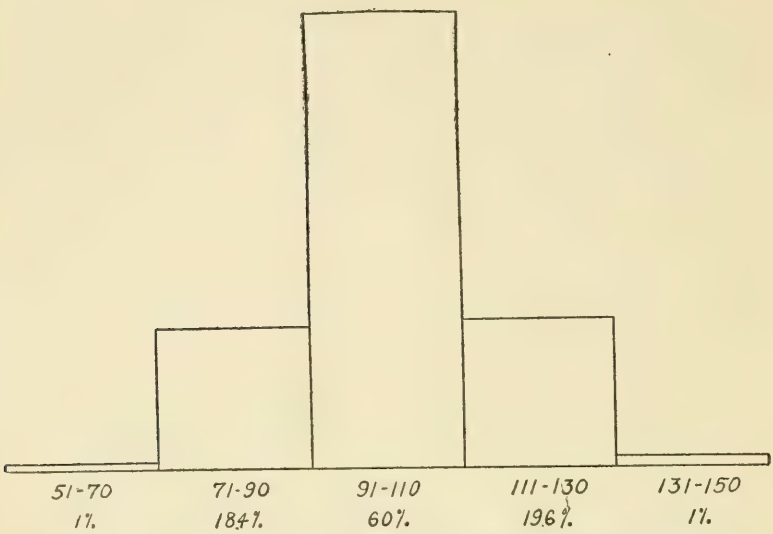
[1] At 15 and 16 there is the additional reason, that growth at this time is probably slowing down sufficiently to impair the validity of the intelligence quotient.

Exception may be taken to this combined distribution on the ground that it fails to take account of possible selection which may have taken place at ages 5, 6, 13, and 14. As shown in Graph 15, however, the distribution for ages 7 to 12 combined is little different from that for 5 to 14 combined.



GRAPH 15. Distribution of intelligence quotients of 554 unselected children 7–12 years of age. (Sexes combined.)

The intelligence quotients may, of course, be grouped in ranges of any desired extent. If the numbers dealt with had been larger, it would have been interesting to group them in ranges of 5 instead of 10. With the numbers available, however, the curves resulting from this method of grouping would be much less regular than when larger groups are used. On the other hand, if ranges above 10 are used for the grouping, the distribution becomes still more symmetrical. This may be seen from Graph 16, which shows the relative sizes of groups contained in ranges of 20, for ages 5 to 14 combined.



GRAPH 16. Distribution of intelligence quotients of 905 unselected children, age 5–14 years. Intelligence quotients grouped in ranges of 20.

Following is a comparison of the observed distribution (Ages 5–14) with that called for by the theoretical "normal" curve of distribution (the Gaussian curve):

TABLE 4

INTELLIGENCE QUOTIENT RANGES

	56–65	66–75	76–85	86–95	96–105	106–115	116–125	126–135	136–145
Obtained	.33	2.3	8.6	20.1	33.9	23.1	9.	2.3	.55
Theoretical	.4	3.1	11.	22.	27.	22.	11.	3.1	.4

The normal nature of the distribution of mental ability is further borne out by the teachers' rankings according to intelligence and school work, which were made according to the form given in the appendix. These rankings were made in a little more than half

the schools and included 489 children, about equally divided between the sexes. The results are shown on page 48.

The distribution shown in the graphs of this chapter are for the boys and girls combined. There are certain sex differences, however, which will be set forth in Chapter III. We may ignore these for the present, since they are not great, and proceed to an examination of the frequency with which various departures from the normal are encountered.

An examination of the distribution for all the ages from 5 to 14, taken together, and for the sexes combined, gave the following significant facts:

TABLE 5

The lowest	1	percent	go	to	70	or	below;	the	highest	1	percent	reach	130	or	above
"	"	2	"	"	73	"	"	"	"	2	"	"	128	"	"
"	"	3	"	"	76	"	"	"	"	3	"	"	125	"	"
"	"	5	"	"	78	"	"	"	"	5	"	"	122	"	"
"	"	10	"	"	85	"	"	"	"	10	"	"	116	"	"
"	"	15	"	"	88	"	"	"	"	15	"	"	113	"	"
"	"	20	"	"	91	"	"	"	"	20	"	"	110	"	"
"	"	25	"	"	92+	"	"	"	"	25	"	"	108	"	"
"	"	33.3	"	"	95	"	"	"	"	33.3	"	"	106+	"	"

A perfectly normal distribution would cause 93 percent of the cases to fall between 76–125, instead of the observed 94.7 percent; and 99.2 percent between 66–135, instead of the observed 99.3 percent.

Or, to put some of these facts of distribution in another form, we may say, speaking approximately:

The child reaching			110	is	equaled	or	excelled	by	20	out of	100
"	"	"	115	"	"	"	"	"	10	" "	100
"	"	"	125	"	"	"	"	"	3	" "	100
"	"	"	130	"	"	"	"	"	1	" "	100

Again, for those whose intelligence quotient is below 100:

The child testing at			90	is	equaled	or	excelled	by	80	out of	100
"	"	"	85	"	"	"	"	"	90	" "	100
"	"	"	75	"	"	"	"	"	97	" "	100
"	"	"	70	"	"	"	"	"	99	" "	100

When we examine the above data, it is difficult to avoid the conclusion that superior intelligence of any given degree occurs with approximately the same frequency as intelligence which is inferior to a corresponding degree. The usual assumption, however, is that extreme degrees of mental deficiency are much more numerous than extreme degrees of mental superiority. As far as intelligence quotients between 60 and 140 are concerned, our figures do not support this assumption.[2] As regards the relative frequency of intelligence quotients below 60 and above 140, the assumption is in all probability valid; for while all idiot and most imbecile children have an intelligence quotient as low as 50, there are extremely few cases of budding genius which reach as high as 150. Indeed, notwithstanding diligent search, the writer has found only a few cases testing above 150, and only two testing as high as 170.[3]

The significance of various intelligence quotients will be dealt with more fully in another chapter. It is

[2] This statement requires some modification in view of the fact that our data were collected entirely from children who were attending public schools. There are, of course, in any community a few children with intelligence quotients between 60 and 80 who are either kept at home or placed in institutions. No investigation seems to have been made which would show what proportion of such children are not in school, but our experience suggests that it is very small. At any rate, relatively few children testing as high as 65 or 70 are sent to an institution until they have first been tried for several years in the schools, usually until well toward adolescence. It is well known that only a small minority with an intelligence quotient of 75 to 80 ever get into an institution for the feeble-minded.

[3] This is not to deny that cases of considerably higher intelligence quotient are to be found. A five-year-old child reported by Miss Langenbeck seems to have tested not far from ten years. If the test can be accepted at its face value, the child, therefore, had an intelligence quotient of about 200. See A study of a five-year-old child, *Pedag. Seminary*, March, 1915, 65–88. More recently Dr. Leta S. Hollingworth has favored us with a copy of a test of an eight-year boy whose intelligence quotient is 190.

already evident, however, that the term "feeble-mindedness" is a matter of arbitrary definition. In one sense it could be said that a child with an intelligence quotient of 85 or 90 is as truly feeble-minded as one testing at 50, only he is mentally feeble to a much less degree. It becomes merely a question of the amount of intelligence necessary to enable one to get along tolerably with his fellows and to keep somewhere in sight of them in the thousand and one kinds of competition in which success depends upon mental ability. The definition of feeble-mindedness, too, is a constantly shifting one. Until recent years the standard was one which would have classed a majority of children having two-thirds intelligence (intelligence quotient 67) in the normal group. Even yet the usual medical standard is no higher than this. The child of moron grade is rarely classified by the physician as "feeble-minded." Social workers, psychologists, and criminologists, however, are constantly meeting facts which would seem to justify the application of the term feeble-minded to many children with three-fourths intelligence (intelligence quotient 75). It is possible that the development of civilization, with its inevitable increase in the complexity of social and industrial life, will raise the standard of mental normality higher still.

But whatever the standard, the number of borderline and debatable cases will probably be greater than the number of those whom all would agree to call feeble-minded. The attempt to classify all children as either definitely feeble-minded or definitely normal involves exactly the same difficulties as we should encounter in trying to classify all adult men as either "normally tall" or "abnormally short," and we may add that

the one attempt is just about as much worth while as the other.

To regard feeble-intelligence as always a disease, which, like small-pox, one either has or does not have, is a view which is contradicted by all we know about the distribution of mental traits. Physicians find special difficulty in freeing themselves from this fallacy, since for them diagnosis consists essentially in determining the presence or absence of definitely pathological conditions.

There is other evidence than that of mental tests to support the hypothesis that intelligence is distributed in the manner indicated by our distribution of intelligence quotients. It has often been shown that a similar distribution results when teachers are asked to classify children into three groups (or five groups) according either to school success or intelligence. Thus, Bobertag had teachers classify 2772 pupils into three groups according as their school work was "unsatisfactory," "satisfactory," or "better than satisfactory." The resulting classification showed 50.8 percent in the middle group, 25.7 percent in the superior group, and 23.5 percent in the inferior group. The numerous other studies which have been made of teachers' marks give similar results, though in some cases the curve shows a more noticeable tendency to be skewed in the direction of superiority.

About half our teachers supplied for each of their children tested the supplementary data called for on the blank shown in the appendix. It will be noted that Question 4 calls for a classification of the children into five groups according to the quality of school work, and Question 5 a similar classification on the basis of the teacher's judgment as to a child's intelli-

gence. The classes were designated in the blank as "very inferior," "inferior," "average," "superior" and "very superior." The resulting distributions are shown herewith.

TABLE 6

PERCENTS RATED

	Very Inferior	Inferior	Average	Superior	Very Superior
School work (503 Children)	5.2	17.9	51.0	22.1	3.8
Teacher's judgment of intelligence (489 Children)	3.4	14.4	55.8	23.1	3.3

If for quality of school work, we combine the two superior groups and combine similarly the two inferior groups, the distribution coincides remarkably with that found by Bobertag.

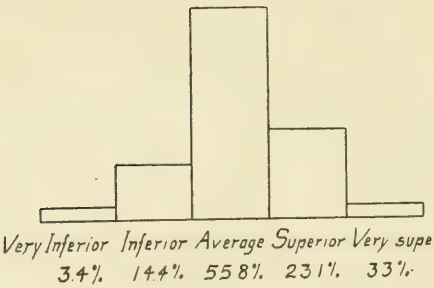
TABLE 7

	Inferior	Average	Superior
Bobertag	23.5	50.8	25.7
Ours	23.1	51.0	25.8

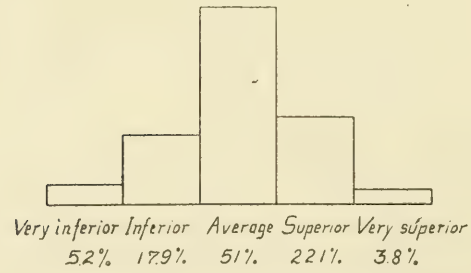
It is interesting to compare the teachers' groupings for intelligence with those for school success, as shown in Graphs 17 and 18.

The "piling up" of the intelligence distribution in the direction toward the "superior" end indicates that teachers are able to judge the degree of a child's intelligence less objectively and therefore less accurately than they judge the quality of his school work. Personal factors are more likely to enter into judgments about intelligence, and in case of uncertainty as to the proper classification there is a natural tendency to give the child the benefit of the doubt, even at the risk of grading him too high.[4]

[4] The relation between the intelligence tests and the teachers' judgments of intelligence is treated in full in Chapter VI.



GRAPH 17. Distribution of teachers' rankings of 489 children according to intelligence.



GRAPH 18. Distribution of teachers' rankings of 503 children according to quality of school work. (Mostly the same children as appear in Graph 17.)

The "Mental Ages" of 62 Normal Adults

The use of the Stanford revision with 30 business men of moderate success and of very limited educational advantages gave a distribution of mental ages differing little from that found with 32 high school pupils pupils from 16 to 20 years of age. The results for both groups are shown in tabular form.

TABLE 8

MENTAL AGE

	13 to 13–11	14 to 14–11	15 to 15–11	16 to 16–11	17 to 17–11	18 to 18–11
Business Men.....	1	6	7	8	6	2
H. S. Pupils.....	0	5	12	10	4	1

If we combine the business men with the high school pupils who are over 16 years of age chronologically, we have the distribution of mental ages shown in Graph 19. It will be noted that the middle part of the graph represents the "mental ages" falling within the range of two years, namely 15 to 17. This range we may designate as the "average adult" level.

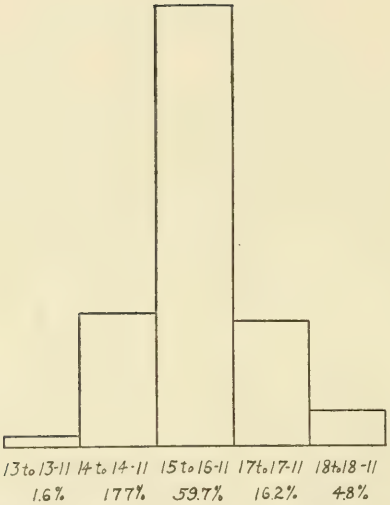
Summary

1. The revised scale gives a median intelligence quotient of approximately 100 when used with unselected children of any age from 5 to 14.

2. The distribution of intelligence quotients for unselected children of each age conforms fairly closely to the Gaussian curve. This holds particularly for our subjects of ages 7 to 13.

3. Since the revised scale yields the same form of distribution of intelligence quotients at each age, it is permissible to combine the intelligence quotients for the different ages from 5 to 13 or 14 into a single surface of distribution.

4. The mental ages found by testing 30 uneducated business men and 32 high school pupils over 16 years of age range from the "inferior adult" level to the "superior adult" level, with the greatest number at "average adult."



GRAPH 19. Mental ages of 62 adults, including 30 business men of little education and 32 high school students over 16 years of age.

CHAPTER III

THE RATE OF GROWTH AND THE VALIDITY OF THE INTELLIGENCE QUOTIENT

The previous chapter showed that for unselected children the distribution of intelligence as measured by the revised scale maintains a certain constancy from 5 to 13 or 14 years, *when the degree of intelligence is expressed in terms of the intelligence quotient.* Any given deviation from the median occurs with much the same frequency at all the ages.

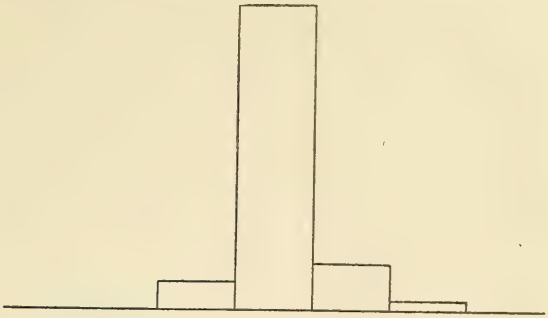
The intelligence of children has usually been estimated, however, in terms of years and months of retardation or acceleration. Binet, while using this method, realized that a year of retardation is less serious with older children than with younger, and accordingly he suggested the rule that while a retardation of 2 years usually means feeble-mindedness in children under ten years of age, older children should not be regarded as feeble-minded unless retarded as much as 3 years. This is obviously crude, but Binet did not suggest any more definite adjustment to allow for the decreasing significance of a given amount of retardation in the upper years. Even this slight adjustment is often ignored by those who use the scale. One person, after testing a large number of juvenile delinquents ranging from 10 to 18 years in age, lumps all the ages together and counts up the number who were retarded 1 year, 2 years, 3 years, 4 years, etc., concluding finally that about 75 percent of the total number were feeble-minded, since that many were retarded 3 years or more. This error appears again and again in the literature of Binet testing. Others, starting from the same erroneous

assumption, have defined "at-age" intelligence as that which is within one year of the child's physical age and have expected to find the number of children testing "at-age" to be the same at all chronological ages.

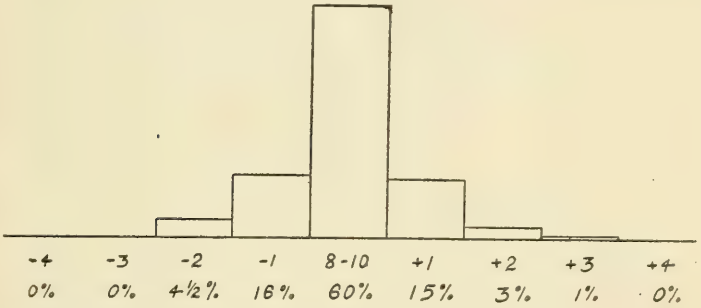
It is obvious, however, from the distribution of intelligence quotients as shown in Chapter II, that a given number of years of retardation can have no definite meaning except in relation to the age of the subject. Whatever the age of a group of non-selected children, approximately the same percent will always be included in any range of the intelligence quotients. As already shown, the middle fifty percent are at all ages included in the range of about 92 or 93 intelligence quotient to 108 or 109 intelligence quotient. Transmuting these values into months, we have for 6-year-olds, 50 percent included in a range of a little less than one year of mental age; while for 12-year-olds, the middle 50 percent range over about twice this distance, or nearly two years. Retardation of two years is about as common at 12 years as retardation of one year at 6; and either is about as common as retardation of a year and a half at 9. That is, the curve of distribution of mental ages becomes progressively flattened, the older the children with which we deal. This is shown in Graphs 20, 21, and 22, which give the distribution of mental ages for children of 6, 9, and 12 years, respectively.

The range including 50 percent of the mental ages increases in a fairly constant ratio from Age 6 to Age 14, as shown in Graph 23.

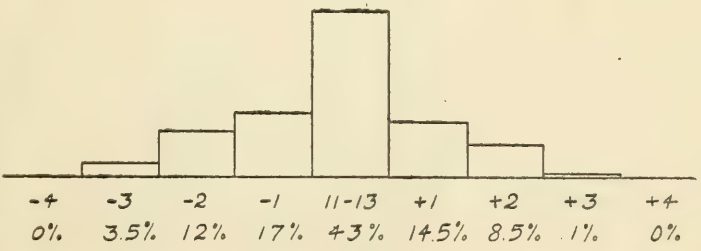
The use of the intelligence quotient as a means of expressing a child's intelligence status is based, of course, on the assumption that the intelligence quo-



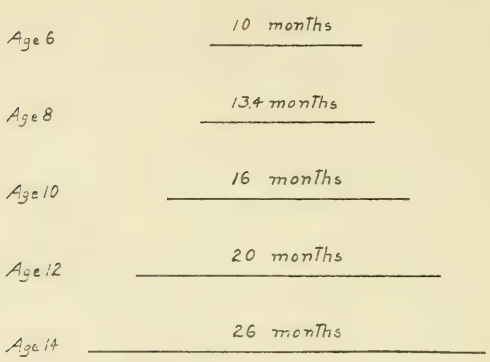
GRAPH 20. Distribution of mental ages of the 117 unselected 6-year-olds.



GRAPH 21. Distribution of mental ages of the 113 unselected 9-year-olds.



GRAPH 22. Distribution of mental ages of the 83 unselected 12-year-olds.



GRAPH 23. Showing range of months including the middle 50% of mental ages at various years.

tient remains practically constant during the years of mental growth; that, for example, the child of five years who tests at 4 (intelligence quotient 80) will at later ages have the mental ages shown in the following figures:

TABLE 9

Physical Age	5	6	7	8	9	10	11	12	13	14
Mental Age	4	4.8	5.6	6.4	7.2	8	8.8	9.6	10.4	11.2

The facts which have already been presented argue in favor of the validity of the intelligence quotient at least for Ages 5 to 13 or 14. It has been shown that the distribution of intelligence quotients for the different years remains essentially the same, and that the distribution of mental ages (in terms of years and months) flattens out in the upper years in approximately the expected ratio; 79 percent test within one year of physical age at 6, 60 percent at 9, and 43 percent at 12. The percents called for at 9 and 12 by a theoretically valid intelligence quotient would be 59.25 and 39.5, granting 79 percent to be correct for

Age 6. The range of mental ages including the middle 50 percent of cases is 10 months at Age 6, and increases in the succeeding years as shown in the accompanying table.

TABLE 10

Age	Range Including Middle 50 Percent	Observed Percentage of Increase Over That of Year 6	Percentage of Increase Called for by a Theoretically Valid Intelligence Quotient
8.....	13.4 months	34	33.3
10.....	16 months	60	67.6
12.....	20 months	100	100

The crucial test of the validity of the intelligence quotient would be to measure the intelligence of the *same children* several times during their period of mental growth. No experiment of this kind appears to have been made on any considerable scale, barring a few repetitions of tests after an interval of only one year. The results of such tests, however, support in a general way the hypothesis that the intelligence quotient of a given child tends to remain constant.[1] The matter has been complicated, however, by the uneven inaccuracy of the Binet scale at different levels.

Repeated tests are being made of a considerable number of children with the Stanford revision, and although the investigation is not complete at this writing, the results of 140 such tests show that as far as the age of 13 or 14, even when the tests are separated by as much as five years, changes of 10 points in 12 are relatively rare. In general, it can be said that the superior children of the first test are found superior in the second, the average remain average, the inferior

[1] See W. Stern: Der Intelligenzquotient als Mass der kindlichen Intelligenz, inbesondere der unternormalen. *Zeitsch. f. Angewandte Psychologie*, 1916, Bd. II, Heft, 1–19. (Argues for the constancy of the intelligence quotient from 7–12.)

remain inferior, the feeble-minded remain feeble-minded, and nearly always in approximately the same degree. The most marked exceptions to this rule are found with the feeble-minded, whose intelligence quotient shows a tendency to decrease considerably.

If future investigations should confirm the validity of the intelligence quotient and its necessary corollaries, the practical consequences would be of the greatest importance. It would mean that, after a mental test consuming no more time than an ordinary medical examination, the psychologist would be able to predict, with some degree of accuracy, the future of the child's mental development. There is nothing else which the average parent would more like to know about the child, and nothing else which would prove of greater value in directing its education.

Whether the intelligence quotient holds even approximately with very young children, or with children much beyond the age of 14, is a question on which the data available afford little light. We are warranted in believing, however, that general intelligence practically ceases to develop by the age of 18 or 20 years. The mental age of high-grade morons appears to change little after the age of 14 or 15 years. With normal children development continues a little longer, though at a decreasing rate. It is practically certain, however, that growth of intelligence comes to a standstill somewhere in the later years of adolescence, and that the cessation is gradual rather than sudden. It is evident, also, that beyond the time when the cessation begins, the intelligence quotient rapidly loses its meaning. The 12-year-old moron with a mental age of 8 years has an intelligence quotient of 67, which

at this period of life probably indicates his mental status fairly well. When 6 years old, the same child probably had a mental age of about 4 years, and when 9 years old, a mental age not far from 6 years. But inasmuch as mental growth slows down rapidly sometime after the age of 14 or 15, the mental age of this subject is unlikely ever to go beyond 10 years. Supposing it to stop at 10, the intelligence quotient, if we continued to use it, would be reduced to 50 at the age of 20 years, to 25 at the age of 40 years, etc. Such a use of the term, of course, would be absurd, since the subject's intelligence is really a constant quantity throughout adult life.

If it could be shown that mental growth continues its earlier rate up to a certain age, say 16, and then stopped quite suddenly, we could continue to use the intelligence quotient with adults of any age by merely ignoring the years beyond 16. That is, all adults, for purpose of reckoning the intelligence quotient, would be regarded as exactly 16 years of age.[2] There are two difficulties, however, with a plan of this kind: (1) mental growth probably does not come to a standstill suddenly; and (2) the time of its cessation is not accurately known.

A practical way to get at the matter is to adopt some hypothesis of this general nature, a quite tentative one, of course and by subjecting it to the pragmatic test of experiment, to see whether it is in harmony with ascertainable facts. If the hypothesis first adopted is unable to satisfy the requirements, it may be altered or replaced by a better. In this way, by

[2] Such a scheme would demand, however, that the upper end of the scale be so framed that the intelligence of superior adults as well as that of superior immature subjects could be expressed in an intelligence quotient above 100.

checking up every step and profiting from our mistakes, we should be able finally to arrive at a solution of the problem which would be correct enough for all practical purposes.

Some of the data which have been presented would seem to justify the assumption, as a tentative working basis, that mental age maintains approximately a fixed ratio to chronological age until the latter has reached about 14 or 15 years, that during the next year or two the ratio diminishes, and that after the chronological age of 17 or 18 years, mental age remains constant. According to this hypothesis, the intelligence quotient would be a proper expression of the intellectual status with subjects as old as 14 or 15 years.

This is the hypothesis which has guided us in the extension of the scale at the upper end. We have at least succeeded in shaping it in such manner that a child, for example, whose mental age at 7 was 8 years (intelligence quotient about 115) will have at 14 a mental age in the neighborhood of 16 (intelligence quotient about 115), with the possibility of further increasing his mental age considerably before growth ceases. Our high school students usually test between 15 and 17 years, as do also Knollin's and Zeidler's business men. College students average slightly higher, as we should expect from the fact that they belong to a selected group.

It will be understood naturally that the numbers expressing such mental ages as 17 years, 18 years, 19 years, etc., have only a conventional value and are not to be interpreted literally. Their use offers a feasible, if arbitrary, method of enabling the superior adolescent or adult to earn a quantitative expression

of his superiority in the tests. Tentatively, we may use the intelligence quotient with normal adults by merely ignoring years of age beyond 16. That is, the adult's chronological age is always, for this purpose, reckoned as 16.

Further trial of our revision by repeating the tests between early and late adolescence, supplemented by tests of different groups of adults, will determine the adequacy of our arrangement. It is not offered as a finished product, but as material for further elaboration and refinement.

Although the intelligence quotient maintains a fairly constant value from rather early in childhood until late in the growing period in the case of children of all grades of intelligence above mental deficiency, it is possible that this constancy may not be maintained with defectives, particularly those of low grade. The child of 4 years who has a mental age of 1 year is an idiot and may never develop higher than a mental level of 2 years, perhaps not so high. His intelligence quotient of 25 at 4 years will gradually diminish, say to 15, at the age when mental maturity is attained in the normal child. We must look to the institutions where low-grade defectives are abundant to supply the facts regarding mental growth in these subjects.

In closing this discussion it may be interesting to point out that the facts presented in this and the preceding chapter are not entirely in harmony with certain wide-spread opinions about the rate of mental growth. The view has often been expressed that intelligence normally develops by alternate leaps and rests. Starting from observations on certain instincts, the doctrine of "nascent stages" has come to be applied to the phenomena of mental development generally

and is now almost a dogma in the literature of child psychology. Researches are rapidly showing, however, that instincts themselves have less a Minerva-birth than we had supposed, a fact which Freudian psychology has demonstrated in the case of the sex instincts. As far as general intelligence is concerned, there is little evidence of periodicity or irregularity. If such periodicity or irregularity occurred, the intelligence quotient of the developing child would be now high, now low, instead of maintaining a fairly constant value.[3]

The facts we have presented offer little hope to the parent who would like to believe that his backward boy who is 6 or 8 years old will "catch up" in the supposed spurt of adolescence. Indeed, the much-talked-of adolescent spurt begins to look like a myth.

These same facts, however, furnish some consolation to the parents of a young genius. It has been the custom for teachers and even for some psychologists to inspire in them a dismal and uneasy fear that such a child is in danger of retrograde development. We hear of great men who in childhood were famous blockheads, and of genius children who became numbskulls! Perhaps it would not be safe to assert that such cases do not occur, for logic teaches that universal negatives are hard to establish. If they do occur, we may suppose that a concrete example will sometime come to light, vouched for by the necessary scientific proof.

---

[3] Baldwin has recently shown that physical growth also proceeds at a nearly uniform rate from 7 to 14 years of age, and that a child's physical status, as expressed in height and weight, maintains a fairly constant position, with reference to norms, from age 6 or 7 through adolescence. Bird T. Baldwin: *Physical Growth and School Progress. Bull. U. S. Bur. of Ed.*, 1914, No. 581.

Summary

1. Retardation or acceleration of any given number of years has no meaning apart from the age of the child, and the method of expressing intelligence status in absolute years only should be abandoned.

2. The distribution of intelligence quotients for the separate years argues strongly in favor of the intelligence quotient as a valid method of expressing a child's development status, at least for the years 5 to 14.

3. Data are presented which indicate that the intelligence quotient of a child of any grade of intelligence remains fairly constant until well into the period of adolescence. Doubt is thrown upon the existence of the supposed "nascent stages," the "adolescent spurt," and other popularly assumed irregularities of mental growth.

CHAPTER IV

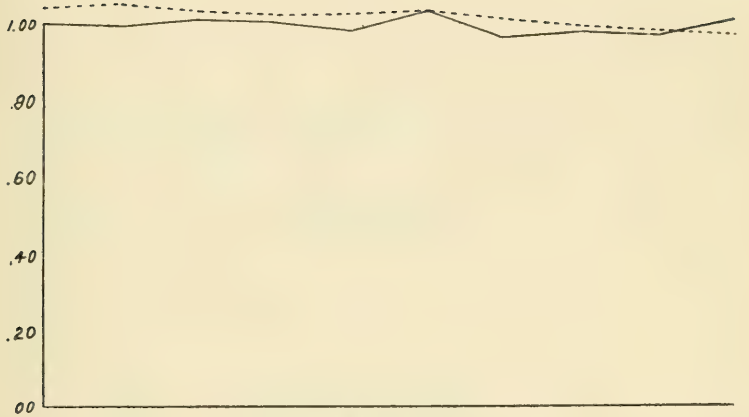
Sex Differences

Our revised scale has been constructed by massing the results from boys and girls, and our discussion of the distribution of intelligence in Chapter II took no account of sex differences. However, when we treat the intelligence quotients of the boys and girls separately, we find a somewhat constant, though rather slight superiority of the girls from Ages 5 to 13, with the exception of Age 10. At 14 years the boys appear to be about as superior to the girls as the girls were superior to the boys at 5. This is shown in Graph 24, which gives the median intelligence quotient for the boys and girls separately at each age from 5 to 14.

Graphs 25–34 show the distribution of intelligence quotients for the sexes separately, when grouped in ranges of ten, 56–65, 66–75, 76–85, 86–95, etc. Because of the small number of boys or girls in any one year, successive ages have been combined for the graphs, as 5–6, 7–8, 9–10, etc.

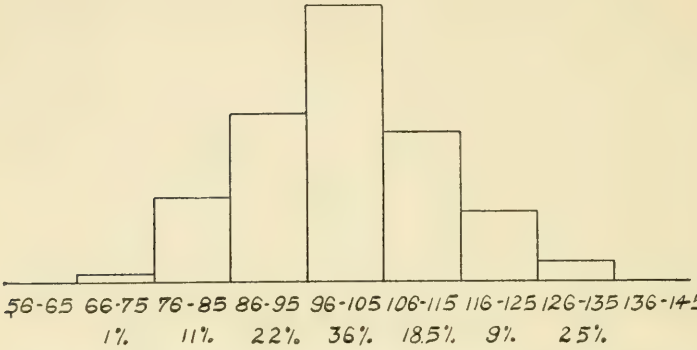
If, now, we combine the intelligence quotients for all ages from 5 to 14, inclusive, we have Graphs 35 and 36 for boys and for girls, respectively.

Comparison of the sexes with regard to the range of intelligence quotient that includes the middle 50 percent shows that half the boys lie between 91–107 and half the girls between 93 and 109. Table 11 shows the frequency of some of the more extreme degrees of variation according to sex.

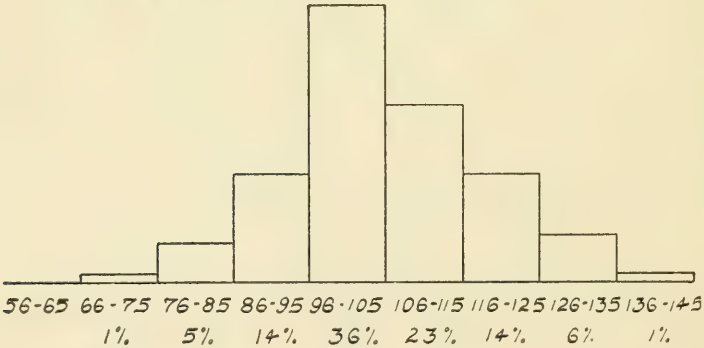


	5	6	7	8	9	10	11	12	13	14
Boys	1.00	.99	1.01	1.00	.98	1.03	.96	.97	.96	1.00
Girls	1.04	1.05	1.03	1.02	1.02	1.03	1.01	.99	.97	.96

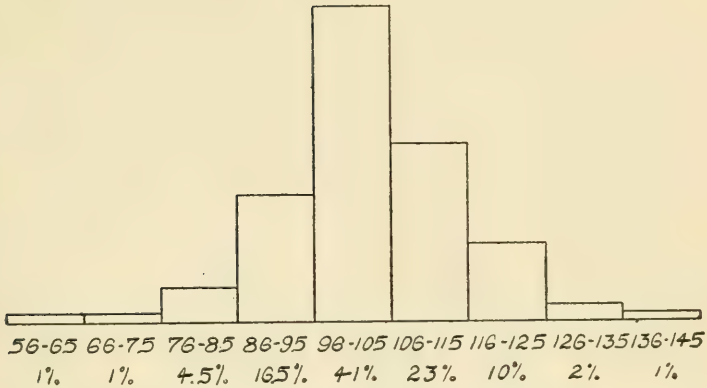
GRAPH 24. Showing median intelligence quotients for boys and girls separately at each age, number of boys, 457; girls, 448.



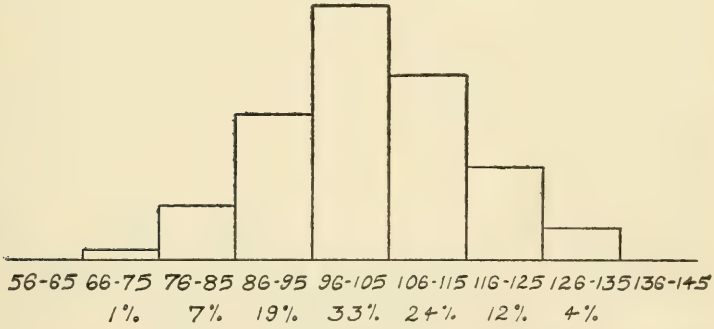
GRAPH 25. Distribution of intelligence quotients of 87 boys, ages 5 and 6 combined.



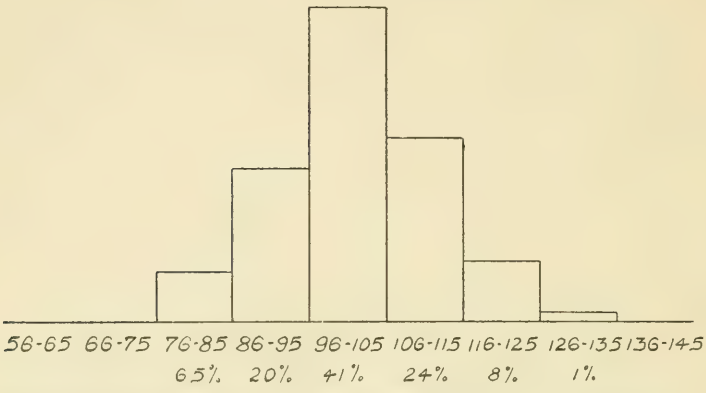
GRAPH 26. Distribution of intelligence quotients of 87 girls, ages 5 and 6 combined.



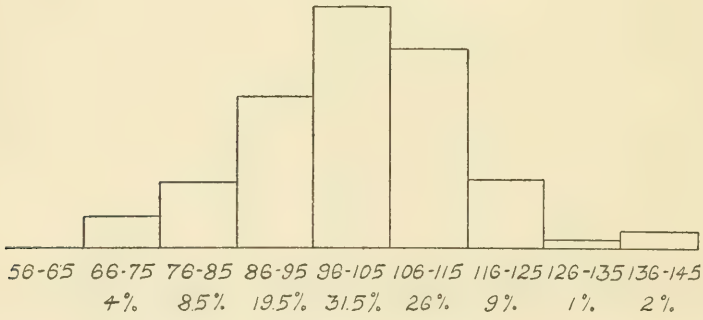
GRAPH 27. Distribution of intelligence quotients of 100 boys, ages 7 and 8 combined.



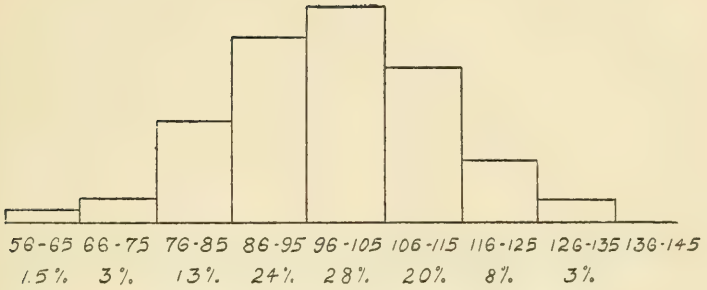
GRAPH 28. Distribution of intelligence quotients of 91 girls, ages 7 and 8 combined.



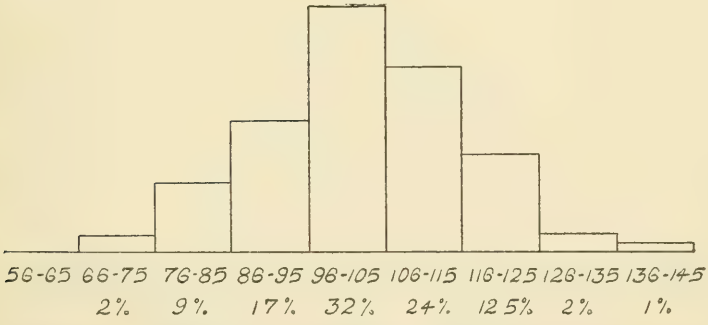
GRAPH 29. Distribution of intelligence quotients of 92 boys, ages 9 and 10 combined.



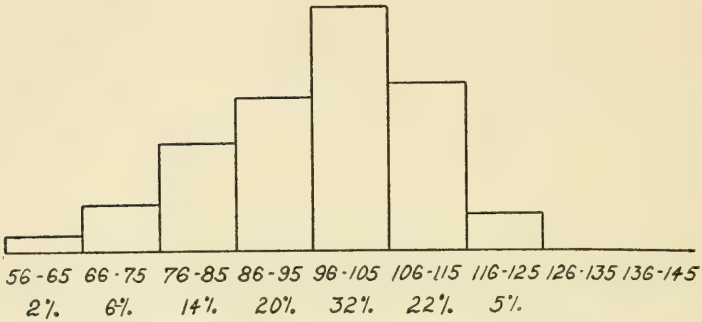
GRAPH 30. Distribution of intelligence quotients of 108 girls, ages 9 and 10 combined.



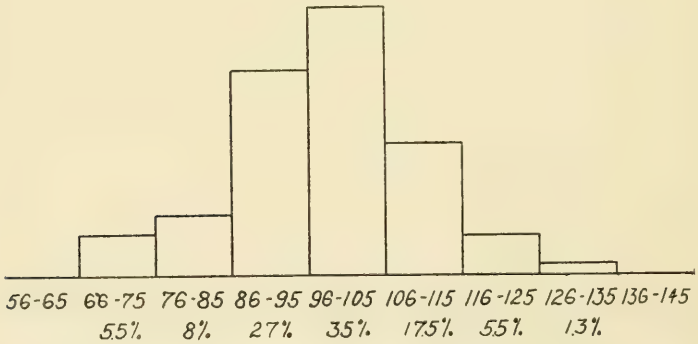
GRAPH 31. Distribution of intelligence quotients of 74 boys, ages 11 and 12 combined.



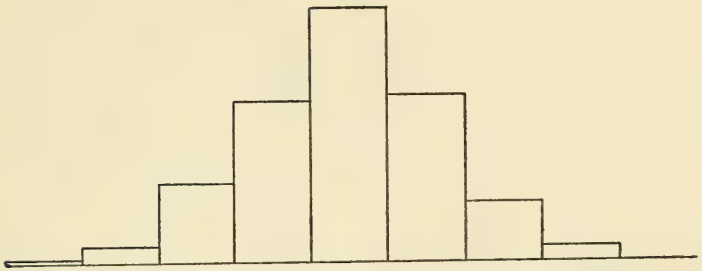
GRAPH 32. Distribution of intelligence quotients of 88 girls, ages 11 and 12 combined.



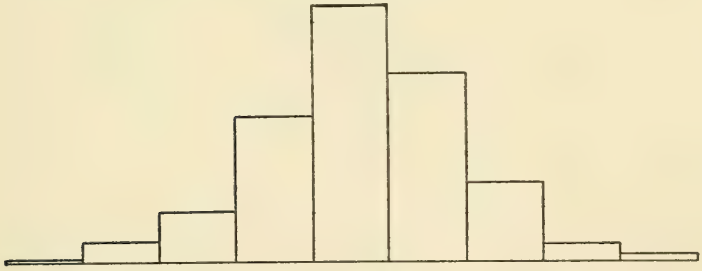
GRAPH 33. Distribution of intelligence quotients of 106 boys, ages 13 and 14 combined.



GRAPH 34. Distribution of intelligence quotients of 74 girls, ages 13 and 14 combined.



GRAPH 35. Distribution of intelligence quotients of 457 boys, 5–14 years of age.



GRAPH 36. Distribution of intelligence quotients of 448 girls, 5–14 years of age.

TABLE 11

DISTRIBUTION OF CERTAIN INTELLIGENCE QUOTIENTS IN 457 BOYS AND 448 GIRLS

	Total 60 or lower		Total 70 or lower		Total 75 or lower		Total 125 or higher		Total 130 or higher		Total 140 or higher	
	No.	%	No.	%	No.	%	No.	%	No.	%	No.	%
Boys...	1	.21%	4	.87%	13	2.83%	11	2.39%	5	1.08%	0	
Girls...	1	.22%	5	1.11%	14	3.12%	17	3.47%	8	1.74%	3	.66%

The facts we have presented indicate that, apart from a slight superiority of the girls from 5 to 12 years, the distribution of intelligence is much the same for the sexes. There is no evidence of any wider range of intelligence among boys, such as has commonly been supposed to exist. The difference, if any exists, seems to be in the other direction. A slightly larger percent of girls than of boys falls to 75 or below, which is the point frequently taken as indicating feeble-mindedness, and a decidedly larger percent of girls reaches as high as 125. The range that includes the middle 50 percent is almost exactly the same in extent for the two sexes. This is all quite contrary to the traditional belief that both feeble-mindedness and exceptionally superior ability are more frequent with boys than with girls.[7]

Although the superiority of the girls is not great in amount, it appears over a long enough period to suggest that it may represent a genuine difference and not some accidental condition of the experiment.

[1] In support of our results we are glad to cite the study of Mrs. Leta Hollingworth: "The frequency of amentia as related to sex," *The Medical Record*, Oct. 25, 1913, which is an analysis of 1000 cases examined in the New York Clearing House for Mental Defectives. See also, by the same author: "Variability as related to sex differences in achievement, *Am. J. Sociology*, Jan., 1914, pp. 510–530; and the comparative variability of the sexes at birth, same *Journal*, Nov., 1914, pp. 335–370.

It was first thought that part of the difference might be accounted for by the fact that two-thirds of the tests were made by women and only one-third by men, but when the intelligence quotients were classified according to the sex of the examiners no such influence was discoverable. There remains the possibility that the superiority of the girls in the tests may be the result of a somewhat more ready facility of the girls in the use of language, or of their greater willingness to respond.

Fortunately the supplementary information furnished by the teachers affords us valuable data as to the genuineness of the sex difference in intelligence quotients. On page 48 was given the grouping of 476 children, boys and girls together, according to *the teachers' estimates of intelligence.* When these estimates are summarized for the boys and girls separately (Table 12), the superiority of girls appears only in one respect: viz., 12, or 4.8 percent of the girls are classified in the "very superior" group, as contrasted with 3, or 1.3 percent of the boys.

TABLE 12

TEACHERS' ESTIMATES OF INTELLIGENCE FOR 229 BOYS AND 247 GIRLS, BY SEX

	Percent Ranked Very Inferior	Inferior	Average	Superior	Very Superior
229 Boys	3.9	14.4	56.7	23.5	1.3
247 Girls	3.2	13.7	57.0	21.0	4.8

Table 12 is for the ages 5 to 14 combined. The teachers' estimates for the ages 13 and 14 were treated separately from the other ages in order to find out whether the tendency of the girls to lose their advantage in intelligence quotient at this point is confirmed or contradicted by the judgment of the teachers. The results, shown in Table 13, bear out the tests,

for while the teachers have judged the intelligence of the girls at earlier ages as fully equal, if not superior to that of the boys, they give the advantage at age 13–14 to the boys. The table shows 19.5 percent of the boys of 13 and 14 years of age classed as below average, as contrasted with 24.5 percent of the girls; on the other hand, 26 percent of the boys are classed as above average, as contrasted with 13.5 percent of the girls.

TABLE 13

TEACHERS' ESTIMATES OF INTELLIGENCE OF BOYS AND GIRLS FOR THE AGES 5–12 AND 13–14

Percent Ranked	Very Inferior	Inferior	Average	Superior	Very Superior
Ages 5–12 Combined.....	Boys 3.2	14.7	57.3	22.9	1.6
	Girls 3.8	11.9	56.1	22.3	5.7
Ages 13–14 Combined..	Boys 6.5	13.0	54.3	26.0	0.0
	Girls 0.0	24.5	62.0	13.5	0.0

In like manner we have compared the boys and girls with reference to *the quality of their school work*, as judged by the teachers. When we classify the judgments for the ages 5–12 separately from those for 13–14, we have Table 14.

TABLE 14

TEACHERS' ESTIMATES OF THE QUALITY OF THE SCHOOL WORK OF BOYS AND GIRLS FOR THE AGES 5–12 AND 13–14

Percent Ranked		Very Inferior	Inferior	Average	Superior	Very Superior
Ages 5–12	Boys.....	3.8	11.1	60.0	23.8	1.1
	Girls.....	3.3	12.6	57.7	20.8	5.3
Ages 13–14	Boys.....	4.1	26.5	45.0	22.5	2.0
	Girls.....	2.4	19.5	53.6	22.0	2.4

Table 14 agrees with the tests in showing a larger number of cases of greatly superior ability in school work among girls than among boys from 5 to 12 years. The data for Ages 13–14 agree with the tests less closely, for, while the boys are less inferior to the

girls than in the ages 5–12, they are still somewhat inferior.

There remains still another means of checking up the evidence of the tests as to the relative intelligence of boys and girls; we can compare their *age-grade distribution in school.* Fortunately, our data enable us to do this for all the children tested. Table 15 gives the age-grade distribution for Ages 7 to 14. Ages 5 and 6 are left out of account because children of this age have not had time to become retarded or accelerated in school, and those above 14 are eliminated because they represent a selected group—the "left-overs."

TABLE 15

AGE-GRADE DISTRIBUTION OF BOYS AND OF GIRLS FOR THE AGES 7 TO 14. (IN PERCENTS)

Age	Grade	I	II	III	IV	V	VI	VII	VIII
7	Boys.....	77.7	22.3						
	Girls.....	76.2	21.5	2.5					
8	Boys.....	22.0	52.0	24.0	2.0				
	Girls.....	27.0	52.5	18.2	2.5				
9	Boys.....	3.9	32.8	46.2	15.5	1.9			
	Girls.....	5.2	26.1	50.8	14.0	3.5			
10	Boys.....		2.8	25.0	44.4	25.0	2.8		
	Girls.....		8.1	22.5	51.0	16.4	2.0		
11	Boys.....		5.3	8.4	31.6	34.3	20.0	0	
	Girls.....		0	9.3	16.4	37.2	35.2	35.2	2.3
12	Boys.....			13.3	10.8	32.4	40.6	2.7	0
	Girls.....			0	9.1	32.0	47.9	9.1	2.5
13	Boys.....			3.4	10.3	15.5	38.0	27.7	5.2
	Girls.....			0	2.6	5.4	35.0	40.5	16.2
14	Boys.....				2.2	2.2	24.5	13.3	57.6
	Girls.....				3.0	0	24.0	39.5	33.3

Inspection of Table 15 will show that the results lack uniformity. At Age 7, the age-grade status of the girls is slightly better; at 8, the boys have a shade the advantage; at 9, the girls; at 10, the boys; at 11, 12 and 13, the girls are much in advance; while at 14, the boys are the first time decidedly ahead. On the whole, we may say that there is little difference for Ages 7, 8, 9, and 10; that for the next three years the girls are much more advanced than the boys; and that at 14 the boys have much the advantage.

Ignoring Year 7, since at this age pupils have had little time to become retarded or accelerated, we are justified in grouping the ages together as follows: Years 8, 9, and 10; years 11, 12, and 13; and finally, Year 14 by itself. Table 16 shows the percent of boys and girls retarded or accelerated 1, 2, 3, or 4 years in each of these age-groups.

TABLE 16

PERCENT OF BOYS AND GIRLS RETARDED OR ACCELERATED IN SCHOOL BY 1, 2, 3, OR 4 GRADES IN VARIOUS AGE-GROUPS

Age	Sex	—4	—3	—2	—1	Normal	+1	+2	+3	+4
Ages 8, 9, 10 combined	Boys...			1.4	27.0	48.1	21.1	2.1		
	Girls...			4.6	25.3	51.3	16.0	2.6		
Ages 11, 12, 13 combined	Boys...	1.5	10.0	12.3	34.6	33.0	8.5			
	Girls...		.8	8.0	27.5	42.0	20.1	1.6		
Age 14	Boys...	2.2	2.2	24.5	13.3	57.8				
	Girls...	3.3		24.2	39.3	33.3				

TABLE 17

PERCENT OF BOYS AND GIRLS RETARDED OR ACCELERATED IN SCHOOL BY 1, 2, 3, OR 4 GRADES, FOR THE AGES 8 TO 14 COMBINED

	—4	—3	—2	—1	Normal	+1	+2	+3	+4
Boys	.95	4.47	9.58	28.11	43.13	12.77	.95		
Girls	.32	.32	8.11	27.59	45.45	15.90	1.94	.32	

Table 17 shows the percent of boys and girls retarded or accelerated 1, 2, 3, or 4 years for all the ages 8–14 combined. It should be emphasized, however, that the facts we want to know are best disclosed in Table 16, and that the evidence goes to show *that the grade progress of our boys and girls differs little up to, and including Age 10, that for the next three years the girls are clearly in advance, and that the reverse is the case at 14. In the main, therefore, the school progress of our subjects agrees with the intelligence tests, with the teachers' estimates of intelligence, and with the teachers' judgments of the quality of the school work, in showing a sex difference which is in favor of the girls before 14, and in favor of the boys thereafter.*

Before accepting this conclusion there is one other factor to be taken into account which might help to explain the apparent superiority of the boys at Age 14. A certain amount of selection has taken place in this age-group. A considerable number of the 14-year-olds have been promoted to the high school, and these are not included in our group of subjects of this age. This has doubtless occurred more often with the girls than the boys, for we have already showed that marked school acceleration occurs much oftener with girls than with boys at the ages 12 and 13 years. At Age

12 there are 9.1 percent of the girls in Grade VII and 2.5 percent in Grade VIII, as compared with only 2.7 percent of the boys in Grade VII and none at all in Grade VIII. Similarly, at Age 13 there are 16.2 percent of the girls in Grade VIII as compared with only 5.2 percent of the boys. If all our 13-year-olds in Grade VIII should be promoted to the high school at the end of the year, the number of 13-year-old girls receiving such promotion would accordingly be more than three times as great as the number of boys. If this situation holds true generally, it can not be safely ignored in making a comparative study of the intelligence of boys and girls at the age of 14. Even at 13, unequal selection appears to have taken place to no small degree, as would appear from the following age-grade distribution (in percents) of 13-year-old boys and girls:

TABLE 18

		III	IV	V	VI	VII	VIII
13 year olds	Boys.....	3.4	10.3	15.5	38	27.7	5.2
	Girls.....		2.6	5.4	35	40.5	16.2

The foregoing distribution would suggest that our 13-year-old boys are almost free from any selection as far as promotion to the high school is concerned, but that probably 5 percent or more of the girls who ought to be present at this age are not in our group. Comparison of the relative number of boys and girls tested at different ages fully confirms this suspicion. Table 19 shows that, while the number of girls found in the grades equals or exceeds the number of boys at every age but one below 13, yet at 13 and 14 the girls make up only about 41 percent of the entire number. This is significant when it is remembered that, in all the schools where testing was done, *all*

the boys and girls below the high school were tested who were within two months of a birthday, whatever grade they may have been in. By the laws of chance the number of boys and girls found at each age ought to have been nearly equal, barring selective influences.

TABLE 19

PERCENT THE GIRLS FORM OF THE ENTIRE NUMBER OF PUPILS IN GRADES BELOW THE HIGH SCHOOL AT EACH AGE

Age	5	6	7	8	9	10	11	12	13	14	15	16
Percent...	50	50	49.5	46	51.3	57.5	56	53	41	41.5	45	28.6

The only possible conclusion seems to be that the apparent superiority of boys at the age of 14, as well as also their diminished inferiority at 13, is due solely to the unequal selection which has taken place at these ages. The results of the tests themselves, the teachers' estimates of intelligence, the teachers' judgments about the quality of school work, and the age-grade distribution, offer four separate and distinct lines of evidence pointing in this direction. The same four lines of inquiry are also in general agreement in showing a distinct, though slight superiority of the girls in the ages below 13.

Unfortunately, most of the studies made with the Binet tests throw little light on sex differences. In only two studies besides our own have the subjects tested been nearly enough non-selected to make such a comparison worth while, namely, in Goddard's tests of 1500 Vineland school children and in Kuhlmann's tests of 1000 Faribault school children. Kuhlmann, however, has not analyzed his data for the sexes separately, and Goddard has done so only to the extent of giving the percent of boys and girls testing 1, 2, 3, 4, or more years above or below age.

Even this is given by Goddard only for all the ages taken together, a procedure which ignores the unequal significance of a given number of years of retardation with children of different ages. However, his data for the ages combined agree with our own in indicating a slight superiority of the girls.

TABLE 20

SEX DIFFERENCES AS SHOWN BY THE BINET TESTS (GODDARD)

Percents	—2 years or more	—1 year	At age	+1 year or more	+2 years or more
Boys.....	18.5	23	34.5	20	4
Girls.....	18.5	17	36.5	23	5

Bloch and Preiss made comparisons for sex differences in their tests of 79 boys and 71 girls aged 7 to 11 years. However, their results, which indicated a marked superiority of the boys, throw no light on the question of real sex differences, for the reason that their subjects had been selected from a large number as being supposedly "average" in intelligence, and we have no means of checking up the effect of this selection.

The only other Binet results thus far published which may be used for comparative purposes are those of Yerkes and Bridges, who present the sex differences found in the use of their Point Scale with 760 grammar school children from the kindergarten to the eighth grade. The school chosen was located in a "medium to poor" district, but otherwise there seems to have been no selection of subjects which could have influenced sex differences except the fact that in this study, as in our own, the 14 and 15-year-olds tested were composed wholly of those who had not progressed in school beyond the eighth grade.

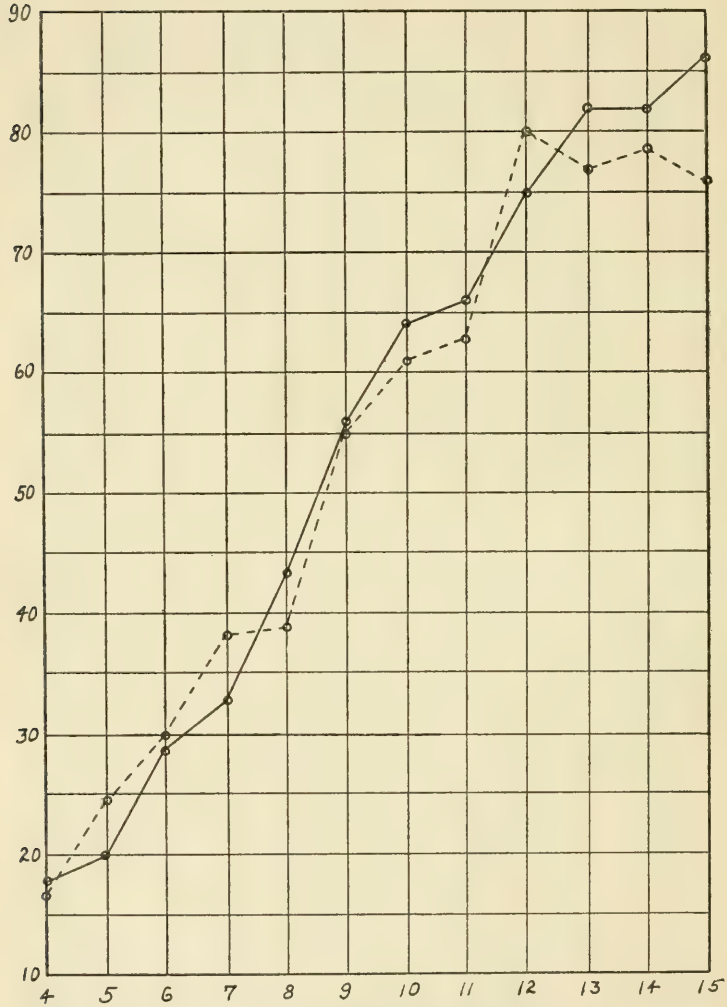
To make comparison easier, Yerkes' and Bridges' curves showing the average scores earned by boys

and girls of different ages are reproduced here. We have chosen the curves obtained by excluding the children of non-English-speaking parents. When the latter were included, the curves crossed and re-crossed so often as to have no clear significance.

It will be noted that the results of Yerkes and Bridges are not altogether in harmony with our own. The two scales agree in that both show a superiority of the girls in the earlier ages and a superiority of the boys at Ages 14–15. The latter, in all probability, has the same cause in both cases, namely, the more frequent elimination of 13, 14, and 15-year-old girls from the grades by promotion. Yerkes and Bridges, however, seem not to have considered this possibility. In the middle ages the results of the two studies are quite in contrast.

In evaluating these somewhat contradictory findings it should be remembered, (1) that the Stanford data are based on more than twice as many children as those entering into the Yerkes-Bridges curve; (2) that all of our children were within two months of a birthday, thus obviating large possible errors likely to result from dividing a small number of children with age differences ranging up to one year into sex groups; and (3) that the number of tests in the Stanford revision is much larger than that included in the Yerkes-Bridges scale, thus reducing the part played by chance.

In conclusion, we may say that the evidence seems to us to point to the existence of a small sex difference in intelligence, which, but for the influence of selection, would probably be in favor of the girls at all ages from 5 to 13 or 14 at least. It should be emphasized, however, that the difference is small, amounting to no



AVERAGE SCORES EARNED BY 468 BOYS AND GIRLS OF DIFFERENT AGES WITH THE YERKES-BRIDGES POINT SCALE

more than 6 percent in terms of intelligence quotient at any age, and at most ages from only 2 percent to 4 percent. In view of the wide distribution of intelligence in each sex (from 50 to 150 intelligence quotient), a difference of 2–4 percent in median intelligence would be practically negligible, even if it were demonstrably genuine. The difference actually found is so small that it might conceivably result from a sex difference in temperamental traits having nothing to do with intelligence, such, for example, as a difference in willingness to attempt a novel task, difference in timidity, or what not. We prefer not to indulge in speculation. At any rate we find no reason to share the opinion voiced by Yerkes and Bridges "that at certain ages serious injustice will be done to individuals by evaluating their scores in the light of norms which do not take account of sex differences."

Finally, the individual tests were examined separately for sex differences. Since the number of our pupils of one sex was ordinarily not larger than 40 to 50 at one age, it was found that the increase in the percent passing in successive years was so irregular as to be very confusing. One way out of this difficulty is to mass together the percents of boys (or girls) passing a test at three separate age levels: the age at which the test appears in the scale, and the adjoining ages above and below. We have not deemed a sex difference worth noting unless it appeared in all of these three successive age levels and to such an extent that the superiority averaged 10 percent for the three ages taken together. This is, of course, an arbitrary basis, but some such plan is necessary to escape the confusion and contradiction engendered by chance variations due to small numbers. As will be seen

from Table 21, the number of tests in which significant sex differences appear, is not large—only 19 out of 58.[2]

TABLE 21

SEX DIFFERENCES IN INDIVIDUAL TESTS

SUPERIORITY OF BOYS		SUPERIORITY OF GIRLS	
Tests in which the superiority of BOYS in three successive years averaged 10 percent or more	Amount of such superiority in percent	Tests in which the superiority of GIRLS in three successive years averaged 10 percent or more	Amount of such superiority in percent
Arithmetical reasoning, XIV	33	Designs from memory, X	19
President and king, XIV	25	Aesthetic compr., V	17
Form-board, IX	20	Ball and field, VIII	16
Fables, XII	15	Giving differences, VII	13
Making change, IX	15	Comprehension, VIII	12
Hands of clock, XIV	12	4 digits backwards, IX	11
Ball and field, XII	11	6 digits, X	10
Similarities, XII	11	7 digits, XIV	10
Induction, XIV	10	Bow-knot, VII	10
		Rhymes, IX	10

Smaller differences were found in favor of the boys in Copying a Diamond (VII), Giving Easy Similarities (VIII), Naming the Months (VIII), and Defining Abstract Words (XII); in favor of the girls in Repeating 5 Digits (VII), Naming the Days of the Week (VII) and Hard Comprehension (XII).

We refrain from extended comment on the list of tests in which sex differences have been found. It will probably agree badly enough with anyone's *a priori* expectations. There are apparent contradictions, also, as well as surprises. We have no theory

[2] The tests of Years 3 and 4, and those of the "average adult" and "superior adult" were left out of this comparison because of insufficient data.

to explain why the girls are superior on the ball and field test of Year VIII (score 2) and the boys on the same test at Year XII(score 3); or why the boys are better in the tests of giving similarities, and the girls in the test of giving differences. Perhaps we should have expected the superiority of the boys in arithmetical reasoning, the form-board, and making change; likewise the superiority of the girls in aesthetic comparison, tying a bow-knot, and repeating digits.

Summary

1. The tests indicate a slight superiority of girls over boys at each age from 5 to 13. The apparent superiority of the boys at 14, however, is probably accounted for by the unequal selection which has taken place in the promotion of pupils to the high school.

2. The small superiority of the girls in the tests probably rests upon a real superiority in intelligence, age for age. At least, this conclusion is supported by the age-grade distribution of the sexes, and by the teachers' rankings according to intelligence and quality of school work.

3. Apart from the small superiority of the girls, the distribution of intelligence shows no significant difference in the sexes. The data offer no support to the wide-spread belief that girls group themselves more closely about the median or that extremes of intelligence are more common among boys.

4. Not many of the individual tests show large sex differences in the percent passing in three consecutive years. In certain of the tests, however, the differences were marked and unexpected.

CHAPTER V

THE RELATION OF INTELLIGENCE TO SOCIAL STATUS

In the use of the Binet scale with different social classes it has generally been found that children who come from superior environment test higher than those who come from homes where the degree of culture is inferior. As already noted, the arrangement of tests in the 1908 Binet scale was based on an examination of about 200 children from one of the poorest quarters of Paris. When the scale thus derived was used by Decroly and Degand in the examination of children from wealthy and cultured homes, in a Brussels private school, it was found that many of the tests were passed two years below the location assigned them by Binet. Jeronutti's tests of 144 better-class children of Rome agree closely with those of Decroly and Degand. The same is true of Madame Wolkowitsch's tests of private-school children at Petrograd. On the other hand, Dr. Anna Schubert's data gathered from 229 *lower-class* children in a Moscow orphanage gave a distribution of mental ages skewed in the opposite direction. In fact, none of Dr. Schubert's children were advanced more than one year, only 27 percent tested "at age," while 75 percent were retarded one year or more and 39 percent two years or more.

Binet, himself, took up the question by having tests made of 54 children who had been classified into four groups by the teachers according to social status. Unfortunately, the school chosen was one which was

not attended by children of the highest social classes, and the number tested was very small. The results failed to show any correlation of mental age with social status. A later comparison by Binet of 30 children attending a poorly situated school with 30 others attending a school in a well-to-do neighborhood of Paris showed a marked difference in favor of the better situated children.

The Breslau experiment, of which a partial account has been given by Stern,[1] indicated that pupils of the *Volksschule* are at the age of 10 years somewhere near the level of mental development which is attained by pupils of the *Vorschule* at 9 years. The *Volksschulen* are attended mainly by children of the laboring and lower business classes and the *Vorschulen* by children of the better classes. Likewise, in comparing children of the upper and lower social classes in English infant schools, Winch found the children of the higher class superior to the other group in a majority of the tests.

Study of the data which we have collated for the individual tests of the scale shows that large differences found by investigators in the percentage of children who pass certain tests may often be accounted for by a difference in the social class of the subjects.

Yerkes and Bridges compared 54 pupils of a better-class school with an equal number attending a poor school. The sexes were represented equally, and the pupils were selected in such a way that a boy or girl of the favored group was matched by a boy or girl of approximately the same age from the unfavored group. The comparison showed that by the Yerkes-

[1] The Psychological Methods of Testing Intelligence, These MONOGRAPHS, No. 13, 1914, pp. 54 ff.

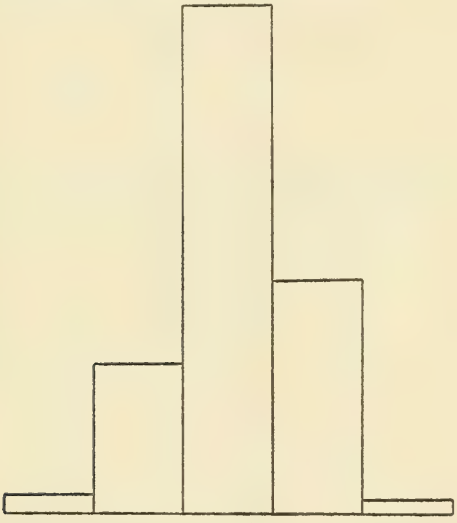
Bridges scale the favored boys averaged 7.7 points higher and the girls 8.4 points higher than did the members of the unfavored group of the same ages. If we compare this difference in points with the age-norms given by Yerkes and Bridges, we find that it represents about a year of difference in mental advancement with children of this age. Binet at one time estimated that social status might make as much as a year and a-half difference in mental age, though in making this statement he seems to have overlooked the fact that a retardation of a year and a-half is not of equal significance in the lower and upper ages.

The tacit assumption which most writers seem to have made in their discussions of such facts as those we have just set forth is that the difference found is due wholly, or at least mainly, to the influence of environment. Meumann believes that the most serious fault of the Binet scale is its failure to take account of the influence of social environment on the ability to pass certain tests. Yerkes and Bridges assert that "it is obviously unfair to judge by the same norm of intelligence two children, the one of whom comes from an excellent home and neighborhood, the other from a medium-to-poor home and neighborhood."

As will be shown presently, we believe that the facts may be more reasonably explained on an entirely different hypothesis. First, however, we will set forth somewhat in detail the Stanford data that bear upon this question.

As stated elsewhere, we were able to secure a classification of 492 of our children according to social class into five groups: "very inferior," "inferior," "average," "superior" and very superior." Although the schools chosen for the tests were on the whole

as nearly average as could be found, it will be readily understood by anyone who is acquainted with the democracy of the American educational system that in almost any small city an "average" school contains some children of every social class. As was expected, therefore, all the social classes were found to be represented in every school. Graph 38 shows the distribution by social class of the 492 children regarding whom the supplementary information was obtained.



GRAPH 38. THE DISTRIBUTION BY SOCIAL CLASS OF 492 CHILDREN OF ALL AGES.

We have next classified these same pupils according to the intelligence quotients resulting from the tests, and for this purpose also we have made use of five groups, as follows:

I Q below 80	I Q 80–89	I Q 90–109	I Q 110–119	I Q 120 or above
"Very inferior"	"Inferior"	"Average"	"Superior"	"Very superior"

This grouping of the intelligence quotients is, of course, arbitrary, but some sort of grouping is necessary, and the one we have employed has the advantage of giving five groups of intelligence quotients which agree roughly in size with the groups according to social class. This correspondence is as follows:

TABLE 22

PERCENT OF PUPILS IN EACH GROUP

I Q	I Q below 80	I Q 80–89	I Q 90–109	I Q 110–119	I Q 120 or above
	6.9	14.5	57.7	15	5.6
Social Status	Very inferior	Inferior	Average	Superior	Very superior
	2	16.6	56.5	20.6	1.6

Table 23 shows where the children in each social group fall with reference to intelligence quotient, and also where those of any intelligence quotient group fall with reference to social class.

TABLE 23

THE RELATION OF INTELLIGENCE TO SOCIAL STATUS

Social Status	I Q					
	Below 80	81–89	90–109	110–119	120 or above	Total
Very Inferior	4	4	3	0	0	11
Inferior	18	15	43	4	0	102
Average	9	43	181	48	11	292
Superior	1	7	59	21	14	80
Very Superior	0	0	3	2	2	7
Total	32	69	289	75	27	492

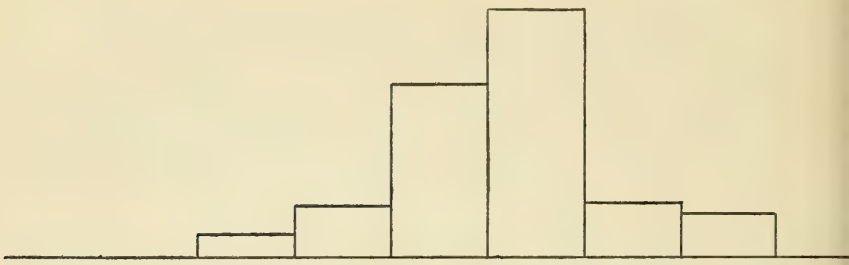
That there is a certain degree of correlation between intelligence quotient and social status is quite evident from a glance at the table. Of the 27 children with an intelligence quotient of 120 or above, not a single one comes from a social class below the average; of the 75 with an intelligence quotient between 110 and 119, only 4 belong to the "inferior" social group, and none to the "very inferior group." Conversely, of the 32 children with an intelligence quotient below 80, only one is classified below the "average" social group. Only the middle intelligence quotient group, 90–109, is represented in all the social classes; and only the "average" and "superior" social group is represented in all the intelligence quotient groups. Application of the Pearson formula to the data in this table gives a correlation of .40 between social status and intelligence quotient.

Another way to express the relationship between intelligence and social status is to compare the median intelligence quotient for the children of each social group, as follows:

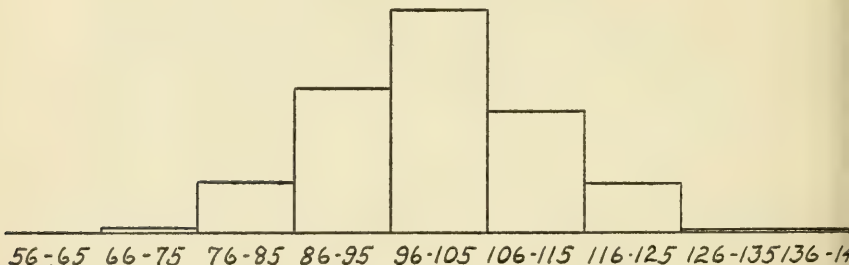
Social Group	Very Inferior	Inferior	Average	Superior	Very Superior
Median I Q	85	93	99.5	107	106

Since only 8 pupils are included in the "very inferior" and only 10 in the "very superior" social groups, the medians for these extreme groups have limited significance. The case is different, however, in the "inferior," "average" and "superior" groups, which include 80, 292, and 102 cases, respectively.

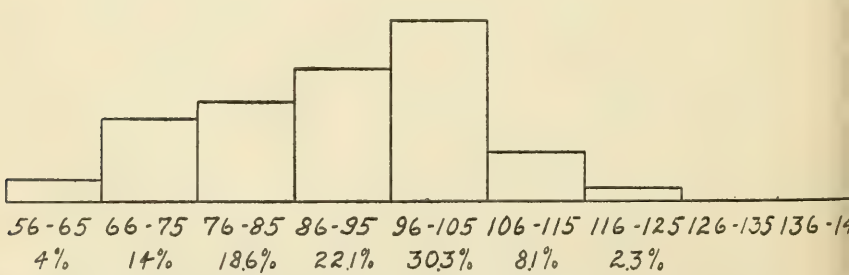
Graphs 39, 40, and 41 show the distribution of intelligence quotients grouped by 10's for the three social classes. In this case the 7 "very inferior" pupils are thrown with the "inferior" class, and the 11 "very superior" pupils with the "superior" class.



#39. Distribution of IQ's of 105 children of "superior" and "very superior" social classes.



#40. Distribution of IQ's of 285 children of "average" social class.



#41. Distribution of IQ's of 86 children of "inferior" and "very inferior" social classes.

The median intelligence quotient of the "average" social group is 99.5, and so practically coincides with the median for all classes taken together. It is significant that the median intelligence quotient of the "inferior" group is 14 points below that of the "superior" group, and that the median intelligence quotient for the "average" group lies approximately mid-way between the two. The difference between the median intelligence quotients of the "inferior" and the "superior" groups means a difference in mental age at the various actual ages as follows:

TABLE 24

MEDIAN DIFFERENCE IN MENTAL AGE

Age		Age	
4	6.7 mo.	10	16.8 mo.
5	8.4 "	11	18.5 "
6	10.1 "	12	20.2 "
7	11.7 "	13	21.8 "
8	13.4 "	14	23.5 "
9	15.1 "	15	25.2 "

Our results, accordingly, agree closely with those of other workers. Our next task is to find the most rational hypothesis which will explain the correlation between social status and intelligence quotient.

The usual assumption has been that the correlation is the artificial product of environmental influences; that the child from a cultured home does better in the tests by reason of his superior home training and because he has had more opportunity to pick up the information which success in the various tests calls for. This explanation has seemed to us from the beginning a most improbable one. Several investigations of the influence of environment on mental traits suggest the conclusion that this influence is much less important than is original endowment in determining the nature

of the traits in question. From an *a-priori* standpoint, the endowment hypothesis explains the correlation between intelligence quotient and social status just as adequately as does the environment hypothesis. To conclude, as Meumann and Yerkes have done, that the demonstration of the existence of such a correlation invalidates the Binet scale as a method of measuring intelligence is to make a gratuitous assumption—an assumption, indeed, which is contradicted by much evidence from investigations bearing on the mental endowment of different social groups. We have thought it worth while, therefore, to sift our data somewhat carefully for evidence on this point.

First, we have compared the social status of the children with the teachers' estimates of intelligence. The tests themselves are brief. Success in some of them, it must be admitted, hinges upon information, the possession of which might conceivably be largely conditioned by home environment. One thinks in this connection of such tests as naming coins, making change, repeating the days of the week and the month of the year, giving definitions, giving the moral of fables, etc. Success in certain others would appear to depend rather too much on facility in the use of language. But the teacher's judgment as to a child's intelligence is based upon months of acquaintance, in this case from half to an entire school year. The teacher has had abundant opportunity to distinguish between real mental ability on the one hand and the accidents of knowledge, or facility in the use of language, on the other. Accordingly, Table 25, which gives the teacher's judgment as to the intelligence of each child in the various social groups, should be of interest.

TABLE 25

THE INTELLIGENCE OF CHILDREN OF VARIOUS SOCIAL CLASSES AS ESTIMATED BY THE TEACHERS

Social Status	Teachers' Estimate of Intelligence					
	Very Inferior	Inferior	Average	Superior	Very Superior	Total
Very Inferior	6	2	3	0	0	11
Inferior	4	31	36	5	1	77
Average	7	34	93	50	4	188
Superior	1	0	40	56	4	101
Very Superior	0	0	0	2	6	8
Total	18	67	172	113	15	385

Casual inspection of this table shows that the judgment of the teachers accords with the evidence from the tests in crediting greater mental ability to the children of superior social status. Not one of the children of the "very superior" social group is ranked below "superior" intelligence, and of the 101 included in the "superior" social group only one falls below "average" intelligence. Conversely, not one of the 11 children of "very inferior" social status ranks above "average" in intelligence, while 6 of them are classified as intellectually "very inferior." By the Pearson formula the correlation between social status and the teachers' estimates of intelligence is .55. This is considerably higher than the .40 correlation found between social status and intelligence quotient.

But children from superior homes are likely to be better dressed, cleaner and more attractive in appearance than children from the poorer homes. Perhaps,

too, they are better behaved, more responsive, and socially more adaptable on account of superior training in the home. It is conceivable that external appearances of this kind, which, all would agree, are in part an expression of home conditions, have deceived the teachers and influenced their ranking of the children according to intelligence.

If this were true, the actual quality of the school work done by the children of various social groups might be expected to afford a corrective for this possible error. School work, because it is more definite and objective, is easier to judge than the complex of mental traits called intelligence. Table 26 shows the quality of the school work, as judged by the teachers, for all the children of the several social groups.

TABLE 26

THE QUALITY OF SCHOOL WORK DONE BY CHILDREN OF THE VARIOUS SOCIAL GROUPS

Social Status	QUALITY OF SCHOOL WORK					
	Very Inferior	Inferior	Average	Superior	Very Superior	Total
Very Inferior	5	4	3	0	0	12
Inferior	10	29	35	6	2	82
Average	9	52	160	60	4	285
Superior	0	4	51	46	4	105
Very Superior	0	0	0	2	6	8
Total	24	89	249	114	16	492

By the Pearson formula the correlation expressed in this table amounts to .47, which is only a little lower than that found for the teachers' estimates and

social status, and somewhat higher than that between social status and intelligence quotient.

We would attach especial importance to this correlation, for the reason that the wide-spread use of pedagogical tests in recent years has demonstrated that the individual differences in subject-proficiency which such tests bring to light among school children represent, in large part, individual differences in native endowment and not the effects of unequal home or school training. Even spelling ability, contrary to common opinion, is largely a function of general intelligence. As Mr. Houser has shown, the correlation between the two ranges from 35 to 71 percent.[2] Individual instruction, "special class" methods, indeed, the concentrated efforts of an entire school system, are unable to wipe out the major differences of this kind in a dozen years. As a rule, the longer the child is in school, the more evident the inferiority of the inferior child becomes. It would hardly be reasonable, therefore, to expect that a little incidental experience and instruction in the home, amounting perhaps in most cases to not more than a few minutes per day, would weigh very heavily against these native differences. Even in good homes, children are likely to learn less from their parents than from their play fellows and nurses.

For further evidence on the relation of school success to social status, we will examine the age-grade distribution of the children in the five social groups. This is shown in Table 27 for the ages 8–16. Children below 8 were omitted from this comparison because

[2] See J. D. Houser: "The relation of ability to general intelligence and to meaning vocabulary. *The Elementary School Journal*, Dec., 1915.

TABLE 27

THE RELATION OF AGE-GRADE DISTRIBUTION TO SOCIAL STATUS

Social Status	Location in the Grades								
	Retarded				In Grade for Age	Advanced			
	4 yr.	3 yr.	2 yr.	1 yr.		1 yr.	2 yr.	3 yr.	Total
Very Inferior	1	1	2	2	2				8
Inferior	3	4	10	27	11	4			59
Average	1	4	21	66	109	21	2		224
Superior			4	16	30	14			64
Very Superior				3	2	1			6
Total	5	9	37	114	154	40	2		321

they have not had time to become much retarded or accelerated in school. Normal progress is here defined as Grade II at Age 8, Grade III at Age 9, etc.

The correlation here is positive, but somewhat smaller than in the case of the intelligence quotient, the teachers' estimates of intelligence, and the quality of the school work. This is what we should expect, knowing that the tendency of schools is to promote children by age rather than by quality of school work or native ability.

Again, if home environment really has any considerable effect upon the intelligence quotient we should expect this effect to become more marked, the longer the influence has continued. That is, the correlation of intelligence quotient with social status should increase with age. We have accordingly

worked out the correlation between intelligence quotient and social status for three separate age-levels: for Years 5, 6, 7 and 8 taken together; for 9, 10, and 11 combined; and for 12, 13, 14, and 15 likewise combined. The coefficients of correlation for these three age-levels were, respectively, .43, .41, and .29. In other words, the longer the supposed powerful influence of home environment is continued, the more independent of it the intelligence quotient becomes. The conclusion indicated is that the home environment, *as environment*, has in all ages of childhood relatively little weight in determining the intelligence quotient.

One other line of argument remains. Anyone who has done much testing knows that if sufficiently large numbers are taken, every degree of intelligence from profound idiocy to very superior ability is represented in every social class. This did not happen to be true in the case of our 500 non-selected children from whom supplementary data were available, but we may be certain that it would have been true if a hundred or a thousand times as many children had been tested. In miscellaneous testing, the data from which are not included in the present study of non-selected children, we have found two children of extraordinarily poor home environment who had an intelligence quotient of 150. The highest we have ever found among children of any class is 170. It is a commonplace that dull and feeble-minded children of all grades of deficiency may be found in any social class. We have tested two feeble-minded children whose fathers were men of substantial reputation for scientific achievement. It goes without saying that in each case the home environment was everything that could have been desired. In each family there are other

children whose intelligence quotients range from 115 to 125. Three children were tested in another family, in which the home conditions were about as wretched as could be imagined. Two of these children had an intelligence quotient between 75 and 85, the third an intelligence quotient of 120. The two former have since shown their inability to do fifth-grade school work by the age of 15 years, while the latter entered the high school at the age of 12. Since the unfavorable home environment did not prevent the superior endowment of the one child from evidencing itself in the tests, we must conclude that the inferior showing of the other two could have not been caused by this same environment.

These are individual cases, and we would not stress them unduly. They do illustrate, however, a most important fact,—that exceptionally superior endowment is discovered by the tests, however unfavorable the home from which it comes, and that inferior mentality can not be overcome by all the advantages of the most cultured home.

Of course, we would not deny all possibility of environmental conditions affecting the result of an intelligence test. On the contrary, we have no doubt that the influence is always present in some degree. What it accounts for in terms of the intelligence quotient, we do not know. That it accounts for the larger and more significant differences seems to us wholly improbable. We have little reason to believe that *ordinary* differences in social environment (apart from heredity)—differences such as those obtaining between the higher and lower classes of children attending approximately the same general type of school in a civilized community—impair the validity of the scale to any great extent.

A crucial experiment would be to take a large number of young children of the lower classes and, after placing them in the most favorable environment obtainable, to compare their later mental development with that of children born into the best homes. No such study, properly safe-guarded, seems to have been made. The study would be quite feasible if carried out with the coöperation of a well-conducted orphanage. Some of the tests which have been made in such institutions indicate that mental subnormality of both high and moderate grades is extremely frequent among children who are placed in these homes. Most, though admittedly not all of them, are children of inferior social classes.

Of 20 orphanage children tested by the writer only 3 were fully normal. The other 17 ranged in intelligence quotient from 75 to 95. Nearly all of these children had been in the orphanage for from two to several years. The orphanage in question is a reasonably good one and affords an environment which is about as stimulating to normal mental development as average home life among the middle classes. The children live in the orphanage and attend an excellent public school in a California village.[3]

After all, does not common observation teach us that, in the main, native qualities of intellect and character, rather than chance, determine the social class to which a family belongs? From what is already known about heredity should we not naturally expect to find the children of well-to-do, cultured, and successful parents better endowed than the children who have been reared in slums and poverty? An affirmative

[3] Additional data will be published shortly on the influence of orphanage life on the intelligence quotient of children who have come from low-grade homes.

answer to the above question is suggested by nearly all the available scientific evidence. The suggestion urged by Meumann and also by Yerkes, that it is unfair to evaluate the intelligence of any child except in terms of the average intelligence of his own social class, is not warranted. It would be just as logical to insist that it is unfair to the dull or feeble-minded child to judge his intelligence with reference to standard intelligence for the mentally normal.

Finally, it should be pointed out that Meumann's strictures on the Binet scale in this connection had their origin in certain discrepancies observed in the results of various investigators, which seemed to him attributable entirely to differences in the social status of the subjects tested. It can be shown, however, that the observed discrepancies may be largely accounted for in other ways. They may have resulted in part from failure to follow the same procedure in giving or scoring the tests. Experience in training a fairly large number of individuals in the correct use of the Binet method leads one to stress this factor as a possible source of large discrepancies in results. Moreover, closer inspection of the discrepancies shows that they are much smaller than Meumann seems to have estimated them. As a criterion of agreement between two workers he uses the relative percentages of children found to test "at age," +1, +2, —1, —2, etc. In certain cases, this criterion has been misleading. It is well known that the original Binet scale was much too easy at the lower end and much too difficult at the upper. Accordingly, whether the mental ages found by a given worker turn out to be predominantly *plus* or predominantly *minus* depends largely on the age of the subjects. If these are young,

the mental ages, by the original Binet scale, will tend to run too high. If they have reached an age which demands the use of the upper tests, the resulting mental ages will be too low. Meumann instances the high mental ages found by Jeronutti in his tests of better-class children in Rome as evidence purely of the influence of *milieu*. However, an examination of Jeronutti's table of results by ages reveals the fact that the excess of *plus* mental ages is present only in the years below 10. Above 10, the mental ages fall predominantly on the *minus* side. Meumann seems to have been similarly misled with reference to the high mental ages found by Madame Wolkowitsch at Petrograd. Since her children were of kindergarten age, we may be sure that their high mental ages resulted in part from the incorrectness of the scale at this point and only in part from the high social status of the children. On the other hand, Meumann instances the excess of *minus* mental ages found in Miss Johnston's tests as an example of the unfavorable influence of low social status on test performance. The fact is that Miss Johnston's excess of *minus* mental ages occurs most noticeably in the years above 9, where the scale is demonstrably too hard. Of her 7-year-olds, only 5 tested *minus* and 24 at or above age.

We would agree with Stern that the remarkable fact is not that minor discrepancies have appeared in the statistics of different workers, but that their results, gathered in France, England, Germany, Russia, Italy, Belgium and diverse parts of the United States, from different classes of children and by methods which have undoubtedly fallen short of the desired uniformity, should agree as closely as they do.

It is quite possible that some of the individual tests of the Binet scale are affected by accidents of environment and training to an extent which largely invalidates them as measures of intelligence. We do not know which tests are included in this class, but research will ultimately disclose their identity. Kuhlmann has shown how unsafe it is to condemn a test off-hand as subject to this or that disturbing influence.[4] The classification of the tests in Meumann's three-fold test series (tests of maturity, tests of endowment, and tests of *milieu*) is based upon inspection, and has little value beyond the program of research which it suggests.[5]

To ascertain the extent to which a test is influenced by environment, apart from endowment, is not easy. An attempt was made to analyze the Stanford data for evidence of this influence on the individual tests, but it was abandoned. Of the 80 to 120 children at each age from 10 to 15 were ordinarily found in each of the two social classes "inferior" and "superior." Such numbers are at best too small to have statistical value, and in this case the matter was further complicated by the presence of large differences due presumably to native endowment. To all appearances, the average child of any social class behaved in the tests like a child of any other class who had the same intelligence level.

---

[4] F. Kuhlmann: The Binet-Simon tests in grading feeble-minded children, *J. of Psycho-Asthenics*, XVI, 1912, pp. 173–193.

[5] See Terman's review of Meumann on the Psychology of Endowment, *J. of Psych.-Asthenics*, 14: 1914–1915; pp. 75–94; 123–134; and 187–199.

Summary

1. The median intelligence quotient for children of the superior social class is about 7 points above, and that of the inferior social class about 7 points below the median intelligence quotient of the average social group. This means that by the age of 14, inferior-class children are about one year below, and superior-class children about one year above, the median mental age for all classes taken together.

2. That the children of the superior social classes do better in the tests is almost certainly due primarily to superior original endowment. This conclusion is supported by five supplementary lines of evidence: (*a*) the teachers' rankings of the children according to intelligence, (*b*) the age-grade progress of the children, (*c*) the quality of the school work, (*d*) the comparison of older and younger children as regards the influence of social environment, (*e*) the study of individual cases of bright and dull children in the same family.

3. In order to facilitate comparison, it is advisable to express the intelligence of children of all social classes in terms of the same objective scale of intelligence. This scale should be based on the median for all classes taken together.

4. Meumann's criticism of the scale with reference to the influence of social environment on the test results was based on insufficient examination of the data, particularly on the failure to take account of the prevailing ages of the children tested in different investigations.

5. In their responses to individual tests, our children of a given social class were not distinguishable from children of the same intelligence level in any other social class.

CHAPTER VI

The Relation of Intelligence to School Success

The degree of school success offers a partial check on the accuracy of the intelligence scale. While we should not expect complete agreement between scholarship and the results of even a perfect test of intelligence, nevertheless a very marked disagreement between the two would suggest some fault either in the tests or in the method of estimating school success.

There are three main indices of the degree of school success which a given child has attained: (1) his advancement in the school grades, (2) the quality of school work he is doing in the grade where he is enrolled, and (3) the extent to which he is regarded by the teacher as intelligent.

At first thought the last point may seem irrelevant. We are taking the term "school success," however, in an inclusive sense. In estimating an individual's success in life we ordinarily take into account not only the objective record of his achievements, but also the impression he has made on his associates and superiors. The latter is a real part of his success. For our present purpose it is important to know whether a child whose intelligence is judged by the teacher to be "inferior" is really capable of doing "superior" school work; or conversely, whether a child judged by the teacher as of "superior" intelligence is likely to do "inferior" school work.

Comparison of Intelligence Quotient with the Quality of the School Work as Judged by the Teachers

Table 28 shows the teachers' judgments as to the quality of school work done by children of the intelligence quotient groups below 80, 80–89, 90–109, 110–119, and 120 or above.

TABLE 28

THE RELATION BETWEEN INTELLIGENCE QUOTIENT AND QUALITY OF SCHOOL WORK (AS JUDGED BY THE TEACHERS)

Quality of School Work	Intelligence quotients					
	Below 80	89–89	90–109	110–119	120–above	Total
Very Inferior	12	5	8	0	0	25
Inferior	9	28	49	6	0	92
Average	8	34	173	32	10	257
Superior	1	5	60	32	15	113
Very Superior	0	0	12	3	2	17
Total	30	72	302	73	27	504

It is evident from the table that a fairly high correlation is present. No child doing "very superior" school work has an intelligence quotient below the middle group 90–109, and no child doing "very inferior" work ranks in intelligence quotient above the middle group. The agreement, however, is far from perfect. The group doing "average" school work contains intelligence quotients all the way from "below 80" to "120 or above." One child with an intelligence quotient below

80 is ranked as doing "superior" school work. The correlation by the Pearson formula is .45.

There are 51 cases out of a total of 504 in which the quality of the school work is two steps removed from the location required for perfect correlation. This is nearly 10 percent of all. An occasional disagreement of one step would naturally be expected, since we know that the quality of school work depends partly on factors other than intelligence, such as health, industry, conscientiousness, quality of teaching, etc. But a disagreement of two steps is serious.

We have examined our data to see if any evidence could be found of the presence of constant factors tending to explain the disagreement between intelligence quotient and quality of school work. In the table are 26 children whose school work is at least two grades better than the intelligence quotient would itself warrant. On looking up the facts about these children, we find that 19 of them are over-age for their grade. Of these, 10 are from two to four years over-age. Of course nothing else is needed to explain the disagreement in these cases. We know it to be true that a 10-year-old child with a mental age of 8 years (intelligence quotient 80) is usually just about able to do school work of average quality in the second grade, or that a 13-year-old with a mental age of 10 years (intelligence quotient 77) can manage to get along fairly well in the third or fourth grade.

Of the other cases of disagreement in this direction, 4 were kindergarten children. Perhaps these should be thrown out altogether, on the ground that the work of the kindergarten fails to bring out clearly differences of intelligence. Of the remaining cases, one girl of 8 years was described by the teacher as "very timid

and sensitive," and it is possible that this caused an unfavorable showing in the tests. Another, also a girl, was described as a child of "wonderfully sweet disposition," that is, the kind of child we are always ready to give the benefit of a doubt. In two other cases there was an additional explanation, namely, an intelligence quotient which was just over the border between two groups. If one of these had tested at 90 instead of 89, and the other at 110 instead of 109, one step of the disagreement would have been eliminated. Accordingly, we may say that of the 26 cases of serious disagreement, 19 are largely accounted for by over-ageness, 4 by the fact that the school attended was a kindergarten which has no formal work, and that in only 2 of the 26 cases was no explanation suggested by the data at hand.

We will now consider the 24 children the quality of whose school work ranked two grades below what the intelligence quotient would lead us to expect. Since over-ageness accounted for nearly two-thirds of the disagreements in the other direction, we will naturally expect to find under-ageness a frequent cause of displacement downward. This is true, but to a less extent than one might have expected. Of the 24 children, 10, or nearly 40 percent, are under-age for the grade they are in. Seven of these, however, are only one year advanced beyond age, a degree of under-ageness which would hardly account for more than half of the observed disagreement in these cases. Four others are kindergarten children and so may be left out of account. In one other case the teacher had contradicted herself. The child in question had an intelligence quotient of 125 but had been ranked only "average" in school work. The teacher's supple-

mentary statement, however, showed that the marks ranged from B to A in every subject except arithmetic, and other statements indicated that the child had unusual talent in composition and in the appreciation of literature. Two cases were sufficiently explained by illness and long absence from school. Another child was described as lazy and incorrigible—traits which would affect the school work unfavorably. In two other cases about half of the disagreement was accounted for by an intelligence quotient just over the dividing line. In the other 4 cases there was nothing in our data which suggested a reason for the inferiority of the school work below apparent intellectual ability, though it is possible that a fuller knowledge of the facts would have cleared up these cases also.

An analysis of the one-step disagreements disclosed the same factors, though here there remained a somewhat larger number for which the data failed to offer a clear explanation. This amount of disagreement, however, is not particularly significant, since it may come about in a great variety of ways having nothing to do with the validity of the tests.

In conclusion we may say that even a two-step disagreement between intelligent quotient and the quality of a child's school work does not in itself argue against the validity of the intelligence test. At least 90 percent of these disagreements are found to be explainable wholly or partly by other facts. It is especially to be emphasized that rarely, if ever, is a child able to do school work, *in the grade where he belongs by age*, more than one degree superior to that which the mental age would lead us to expect. On the other hand, it more often occurs that the quality of

TABLE 29

GRADE DISTRIBUTION OF 676 CHILDREN BY MENTAL AGE

Mental Age	Grade Attended											Total
	I	II	III	IV	V	VI	VII	VIII	HS I	HS II	HS III	
8	25 25.5%	55 56.6%	18 18.4%									98
9	4 4%	24 24.5%	48 49%	19 19.4%	1 1%	1 1%	1 1%					98
10		4 3.8%	30 28.5%	49 46.6%	15 14.2%	6 5.7%	1 .9%					105
11			6 7%	20 23%	38 44.6%	17 20%	3 3.5%	1 1.2%				85
12			1 1%	8 8.3%	19 19.8%	40 41.1%	16 16.6%	12 12.5%				96
13				2 2.6%	7 9%	29 37.1%	19 24.3%	21 27%				78
14			1 1.5%		4 6%	21 31%	16 23.5%	26 38.2%				68
15						5 14%	10 28%	21 58.5%				36
16						2 16.7%	1 8.3%	9 75%				12
												676

the school work drops considerably below the level normal to the intelligence quotient in question. The chief causes are ill-health, irregularity of attendance, or the possession of moral or volitional traits unfavorable to school success.

Correlation Between Intelligence Quotient and Grade Progress

We have made this comparison for the entire number of subjects, but since there is little opportunity for children below 8 years to become retarded we have included in the following table only those with a mental age of 8 years or more. Grade II is regarded normal for Mental Age 8, Grade III for Mental Age 9, etc. The 8-year mental-age group includes all mental ages from 7 years, 7 months, to 8 years, 6 months, and so on.

The table of grade distribution by mental age shows that nine-year intelligence is found all the way from Grade I to Grade VII, inclusive; ten-year intelligence from Grade II to Grade VII, etc. Twelve-year intelligence, which here ranges from Grade III to Grade VIII, would doubtless have been found in the high school also, if tests had been made there in any considerable number.

Table 30 shows the number and percent who, according to mental age, are retarded or accelerated 1, 2, 3 or 4 years. The table includes only the Mental Ages 8 to 16, inclusive.

Table 30 is given to facilitate comparison with the data of others. It should be emphasized, however, that the method of expressing the degree and amount of retardation and advancement in years for children of all ages taken together is misleading.

TABLE 30

AMOUNT OF ACCELERATION AND RETARDATION IN THE GRADES WHEN MENTAL AGE IS TAKEN FOR THE BASIS

	Grade Below Mental Age					Grade Above Mental Age				
	4 Yrs. or More	3 Yrs.	2 Yrs.	1 Yr.	Normal Grade for Mental Age	1 Yr.	2 Yrs.	3 Yrs.	4 Yrs.	Total
Number	4	13	69	184	275	106	22	3	1	676
Percent	.5	1.9	10.2	27.2	40.6	15.6	3.2	.4	.1	

It overlooks the fact that a given amount of retardation or acceleration is not equally significant at the different ages. Here, as in the case of mental age, a deviation of one year at the age of 8 is as serious as a deviation of two years at the age of 16. Obviously, when we measure age-grade progress of school children in terms of years of deviation from normal grade we are using a unit of measure which has no fixed value. Important as this is in the statistical treatment of the retardation problem, it has been consistently ignored in all of the very numerous studies of age-grade progress, from the pioneer work of Ayres on down. As a result, all of these studies are misleading, particularly in the comparisons made as to the "amount" of retardation or acceleration at the various ages and in the various grades.

Reverting to Table 29, it may be pointed out that, after the age of 7 or 8 years, misplacement by one grade is not especially significant, as that could easily happen from any one of a number of causes, such as early or late entrance, illness, a little more or a little less than average industry, etc. But in 112 cases, or nearly 16 percent of all, there is a misplacement of

two grades or more. Eighty-five of these, or 12½ per-cent of all, are cases of grade retardation below mental age; 26, or nearly 4 percent of all, represent grade acceleration beyond mental age. It is interesting to note that school retardation of 2 years or more (reckoned on the mental-age basis) is about three times as common as acceleration of 2 years or more. On the basis of chronological age, the proportion of grade acceleration to grade retardation is even less than this.

Our present task, however, is to find an explanation of the rather surprising disagreement between grade progress and mental age as determined by the scale. Taking up first the 26 children whose grade status is two or more years ahead of their mental age, we find that 19 of these are chronologically over-age for their grade. Ten of the 19 are from two to four years over-age. In other words, those who are accelerated in school on the basis of mental age are usually retarded on the basis of chronological age, certainly an interesting and instructive paradox. The explanation, however, is obvious. The school tends to promote children by age rather than ability, and although the very dull are allowed to become somewhat retarded, this retardation is ordinarily less than would be warranted by their actual mental retardation. For example, there are six children of Mental Age 10 in the sixth grade. Two of these are 14 years of age, two are 15, and one is 16. Of the two children of Mental Age 11 in the eighth grade, one is 17 years old, three are 16, and five are 15. Only two are normal age for the grade.

Turning now to the 85 children who are *retarded* two or more grades below the norm for their *mental age*, we find that 23 percent are, on the chronological

age basis, actually accelerated, and that over half of the remainder are in the grade where they belong by chronological age. Only 8 percent of those who are retarded two years or more according to mental age are retarded as much as two years by chronological age. This confirms the suspicion that promotion is largely governed by chronological age and helps to explain why children of any given mental age are distributed over such a wide range of grades. There are, of course, other factors which sometimes cause children to be enrolled in grades too low for their mental age, e. g., irregularity of attendance, illness, and lack of industry. Unfortunately, our blank for supplementary data did not call specifically for information on these points.

Tables 29 and 30 gave the school progress of the children on the basis of mental age. Tables 31 and 32 give it on the basis of chronological age. Ages below 8 are disregarded, since at this time retardation and acceleration have had little opportunity to occur.

Comparison of Tables 29 and 31 reveals the striking fact that, on the whole, the grade location of school children does not fit their mental age much better than it fits their chronological age. Except in the upper years, children of a given mental age are scattered over nearly as wide a range of grades as children of the corresponding chronological age. Plainly, the efforts made at school grading fail to give groups of children of homogeneous mental ability.

That this is largely due to the incorrect grading of children of inferior and superior intelligence is easily shown by taking those whose intelligence quotient is practically normal, say between 96 and 105, and finding how these distribute themselves in the grades.

TABLE 31

AGE-GRADE DISTRIBUTION OF CHILDREN ABOVE 7 YEARS OF AGE (BY CHRONOLOGICAL AGE)

Chron. Age		Grade I	II	III	IV	V	VI	VII	VIII	Total
8	Number	23	49	20	2					94
	Percent	24.5	52.3	21.2	1.2					
9	Number	5	32	53	16	3				109
	Percent	4.5	29.3	48.4	14.6	2.7				
10	Number		5	20	41	17	2			85
	Percent		6	23.5	48.2	20	2.3			
11	Number		2	7	18	28	22	1		78
	Percent		2.6	9	23	36	28.2	1.2		
12	Number			5	8	26	36	5	1	81
	Percent			6.2	10	32	44.4	6.2	1.2	
13	Number			2	7	11	35	31	9	95
	Percent			2	7.4	11.6	37	32.8	9.5	
14	Number				2	1	19	19	37	78
	Percent				2.6	1.3	24.3	24.3	47.5	
15	Number				1		2	9	35	47
	Percent				2.1		4.2	19.1	74.5	
16	Number						1	1	9	11
	Percent						9	9	8.2	
	Total									678

TABLE 32

NUMBER AND PERCENT OF CHILDREN RETARDED OR ACCELERATED 1, 2, 3, OR 4 YEARS FOR THE AGES 8–14 COMBINED

	Retarded				In Grade for Age	Accelerated			Total
	4	3	2	1		1	2	3	
Number	4	15	55	173	275	89	9	1	620
Percent	.6	2.4	8.8	27.9	44.3	14.3	1.4	.1	100

This method gives the correlation relatively freed from the constant tendency of teachers to over-promote the dull and under-promote the superior children. Table 33 gives this distribution.

TABLE 33

GRADE DISTRIBUTION BY CHRONOLOGICAL AGE OF CHILDREN WITH INTELLIGENCE QUOTIENT BETWEEN 96 AND 105

Chron. Age	Grade								
	I	II	III	IV	V	VI	VII	VIII	Total
8	9	25	4						38
9		10	25	3					38
10			10	19	3				32
11			2	5	17	2			26
12				1	9	10			20
13					1	15	14	1	31
14					1	6	8	15	30
15							4	7	11
16								1	1
Total	31	42	41	28	31	33	26	24	227

Of the 227 children appearing in the above table, only 4 who are below the age of 14 are more than one grade removed from the place where they belong by chronological age. All the two-grade displacements are in the direction of retardation. That is, the child with an intelligence quotient between 96–105 is never found (in our data) two grades advanced in school; and the chances are about 50 to 1 that if he is under 14 years of age and tests between 96–105 he will not be as much as two years retarded. (Of 198 children with ages 8–13, 4 are retarded two years.) At ages 14 and 15 selection has taken place and the proportion of retardation is naturally much larger.

Another interesting comparison may be made by taking the extreme intelligence quotients and finding the location in the grades for the exceptionally dull and exceptionally bright children of each chronological age. We have done this for the intelligence quotients above 120 and below 80. The results are shown in Tables 34 and 35.

Of the 68 children appearing in Table 34, full supplementary information is available regarding 34. Of these, not one is doing less than "average" work in the grade attended, while 23 are graded as doing either "superior" or "very superior" school work. Of the 8 who are advanced two grades beyond chronological age, we have supplementary information regarding 5. Of these 5, every one is graded as doing either "superior" or "very superior" work, and every one is ranked by the teacher as either "superior" or "very superior" in intelligence.

It will be noted that of the 54 children 7 years old or above, 15 are in the grade where they belong by chronological age, while 3 are even retarded one year

TABLE 34

AGE-GRADE DISTRIBUTION OF CHILDREN WITH INTELLIGENCE QUOTIENT 120 OR ABOVE

Chron. Age	Grade									
	Kgn.	I	II	III	IV	V	VI	VII	VIII	Total
6	2	11	1							14
7		6	3	2						11
8			1	7	2	1				11
9					5	1				6
10				1	3	3	1			8
11					1	1	6			8
12							3	3		6
13							1	1	2	4
14										
15										
16										
Total	2	17	5	10	11	6	11	4	2	68

by chronological age. *That is, 18, or one-third of all those 7 years old or older having an intelligence quotient of 120 or above fail to reap any advantage (as far as promotion is concerned) from their very superior intelligence.* They are all doing "very superior" to "average" school work and would doubtless continue the same record if accorded the extra promotions warranted by their intelligence quotient. The reluctance of teachers to give such promotions is probably due both to inertia and to an unwillingness to part with exceptionally satisfactory pupils.

Of the 42 children appearing in the above table, all of whom have between two-thirds and four-fifths intelligence (intelligence quotient 65 to 80), only two are in the grade where they belong by chronological age. Both of these were doing "very inferior" school work and neither was promoted the following year. Six of the 42 are only one year retarded. Supplementary data are available for only four of the six. Two of these four are doing "very inferior" work, two "inferior" work. Of the 18 who are retarded two years, supplementary data are available for 11, four of whom are said to be doing "average" work, four "inferior" work, and three "very inferior" work. Of the 16 retarded three years or more, we have supplementary data for 10, three of whom are doing "average" work, four "inferior," and three "very inferior." It is interesting to note that two of the three who are doing "average" work are four years retarded: one is 13 years old and in the third grade, the other is 14 years old and in the fourth grade. This is really what we should expect of high-grade feeble-minded children of 13 and 14 years.

TABLE 35

AGE-GRADE DISTRIBUTION OF CHILDREN WITH INTELLIGENCE QUOTIENTS BELOW 80*

Chron. Age	Grade								
	I	II	III	IV	V	VI	VII	VIII	Total
8	2								2
9	3	2							5
10		4							4
11		1	2						3
12			4	2					6
13			1	5	2	1	2		11
14				1		2			3
15						1	1	1	3
16							1	2	3
17							1	1	2
Total	5	7	7	8	2	4	5	4	42

* (35 of the 42 have intelligence quotients between 70–79.)

The foregoing is suggestive because indicative of what three-quarter intelligence can do. A child of this degree of deficiency is usually two to four years below grade for his age, and his work is usually "inferior" or "very inferior." Rarely is he found in the grade where he belongs by chronological age and he never does better than "inferior" work there.

We learn less from Table 34 of what pupils of intelligence quotient 125 can do than we do from Table 35 of what pupils of intelligence quotient 75 can do. The reason is that the school does not often give the superior child a chance to work up to his proper level of performance. Compared to their possibilities, children of exceptionally superior intelligence are usually retarded, just as we found exceptionally inferior children almost always above the grade where they belong by mental age. Most of the apparently much-retarded children are really accelerates; many of the exceptionally accelerated children are really retardates.

TABLE 36

CORRELATION BETWEEN INTELLIGENCE QUOTIENTS AND INTELLIGENCE ESTIMATED BY THE TEACHERS

Teachers' Estimates	I Q					
	Below 80	80–89	90–109	110–119	120 and up	Total
Very Superior	0	2	7	4	3	16
Superior	2	3	60	29	17	111
Average	10	37	184	39	7	277
Inferior	11	22	34	0	0	67
Very Inferior	8	4	6	0	0	18
Total	31	68	291	72	27	389

By the Pearson formula the correlation contained in Table 36 is 0.48. This is about what others have found, and is both large enough and small enough to be significant. That it is moderately high in so far corroborates the tests: that it is not higher means that either the teachers or the tests have made a good many mistakes.

We note first that 24 children in the above table are placed two steps higher in the teachers' estimates than the intelligence quotient would suggest, while only 13 are displaced two steps downward. This discrepancy would indicate that there is probably some factor causing teachers to overestimate the intelligence of those whose test performance is low. On looking into the matter we find that of the 24 children misplaced upward by the teachers' estimates, 14 are from two to four years over-age for their grade. These cases are, therefore, sufficiently explained. It is the teachers who were at fault, not the scale. In judging the intelligence of these children they forgot to make allowance for the over-ageness. Finding them about on a par in intellectual maturity with other children of their classes, they judged them equally intelligent.

Of the remaining 10 children in this group, three were in the kindergarten, where the teacher has little opportunity to form an opinion as to a child's intelligence. In another case, that of a boy with an intelligence quotient below 80 who was ranked "average," the teacher had contradicted her own estimate by adding an explanatory note which made it clear that the boy was probably a borderline case or even feeble-minded, though possessed of some ability to profit by drill work suited to children a year or two below his age. In four other cases the intelligence quotient was

just over the dividing line, making the disagreement between it and the teacher's estimate appear almost twice as great as it really was. In only two of the 24 cases was there no information at hand that would explain all, or nearly all of the disagreement.

Of the 13 who were displaced two steps downward in the teachers' estimates, we find that five were from one to two years under-age for their grade. Their intelligence had accordingly been judged by a standard which was unfair to them; that is, by a standard based upon the average intelligence of older children. Two were kindergarten children. In another case the teacher, after ranking the child in the "very inferior" group, added a note saying that the child was very deaf and that this might account for the apparent stupidity. The test gave this child an intelligence quotient of 95, which was probably not far from correct. Half the disagreement could be accounted for in two other cases by the presence of the intelligence quotient near the dividing line. This leaves 3 cases of two-step downward displacement still unexplained, though we are inclined to suspect that if more facts were available, these, too, could have been cleared up.

Similar reasons appear to account for approximately half of the one-step disagreements. When all the explained disagreements were eliminated from Table 9, the correlation rose from .48 to .71.

Another way to get at the degree of agreement between intelligence quotients and the teachers' estimates (as the latter would be if freed from the constant error due to neglect of age differences) is to compute the correlation separately *for those children who are in the grade where they belong by chronological age.* When this is done the coefficient of correlation, as may be found from Table 37, rises from .48 to .57.

TABLE 37

Teacher's Estimates	I Q					
	Below 80	80–89	90–109	110–119	120 or above	Total
Very superior	0	0	1	3	2	6
Superior	0	0	27	20	3	50
Average	0	6	80	16	1	103
Inferior	0	4	12	0	0	16
Very inferior	1	3	2	0	0	6
Total	1	13	122	39	6	181

Still another method of showing how strongly teachers tend to base their estimate of a child's intelligence upon the quality of his school work, to the neglect of age differences, is to take their classification of a group of children according to intelligence, and their classification of the same children according to school work, and ascertain the degree of correlation between the two groupings. We have done this, with the result shown in Table 38.

TABLE 38

CORRELATION BETWEEN THE TEACHERS' GROUPINGS ACCORDING TO INTELLIGENCE AND ACCORDING TO QUALITY OF SCHOOL WORK

Teachers' Classification according to School Work	Teachers' Classification According to Intelligence					
	Very Inferior	Inferior	Average	Superior	Very Superior	Total
Very Superior	0	0	0	3	12	15
Superior	0	0	16	83	6	105
Average	0	12	212	22	0	246
Inferior	4	45	35	1	0	85
Very Inferior	13	11	2	0	0	26
Total	17	68	265	109	18	477

The correlation is .82 and would probably have been still higher if the supplementary form filled out by the teachers had not contained the specific instruction to estimate the intelligence of a child "*in comparison with other children of the same age.*" In spite of this injunction, they have obviously ignored age differences and estimated intelligence chiefly on the quality of the child's school work in the grade where he happened to be. They have failed to realize that quality of school work is no index of intelligence unless age is taken into account. The question should be, of course, not: "is this child doing his school work well?" but rather: "in what school grade should a child of this age be doing satisfactory work?" A high-grade imbecile may do average work in the first grade and a high-grade moron average work in the third or fourth grade, *provided only they are sufficiently over-age for the grade in question.*

Our experience in testing children for segregation in special classes has time and again brought this peculiar fallacy of teachers to our attention. We have often found one or more feeble-minded children in a class after the teacher had confidently asserted that there was not a single exceptionally dull child present. In every case where there has been opportunity to follow the later school progress of such a child the substantial accuracy of the mental test has been confirmed.

The following are typical examples of the neglect of teachers to take the age factor into account when estimating the intelligence of the child over-age for his grade:

A. R. Boy, age 17, mental age 11, sixth grade, school work "nearly average," teacher's estimate of intelligence "average." Test plainly shows this child to be a high-grade moron, or borderliner at best. Had attended school regularly 11 years and had made 6 grades. Teacher had compared child with his 12-year-old classmates.

H. A. Boy, age 13, mental age 9–6, low fourth grade, school work "inferior," teacher's estimate of intelligence "average." The teacher

blamed the inferior quality of school work to "bad home environment." As a matter of fact, the boy's father is feeble-minded and the normality of the mother is questionable. An older brother is in a reform school. We are perfectly safe in predicting that this boy will not complete the eighth grade, even if he attends school till he is 21 years of age.

F. I. Boy, age 12–11, mental age 9–4, third grade, school work "average," teacher's estimate of intelligence "average," social environment "average," health good and attendance regular. Intelligence and school success are what we should expect of an average 9-year old.

D. A. Boy, age 12, mental age 9–2, third grade, school work "inferior," teacher's estimate of intelligence "average." Teacher imputes inferior school work to "absence from school and lack of interest in books"! We have yet to find a child of 75 intelligence quotient who *was* particularly interested in books or enthusiastic about school.

C. U. Girl, age 10, mental age 7–8, second grade, school work "average," teacher's estimate of intelligence "average." Teacher blames adenoids and bad teeth for retardation. No doubt of child's mental deficiency.

P. I. Girl, age 8–10, mental age 6–7, has been in first grade two years and a half, school work "average," teacher's estimate of intelligence "average." The mother and one brother of this girl are feeble-minded.

H. O. Girl, age 7–10, mental age 5–2, first grade for 2 years, school work "inferior," teacher's estimate of intelligence "average." The teacher, nevertheless, adds: "This child is not normal, but her ability to respond to drill shows that she has intelligence." It is true that even feeble-minded children of 5-year intelligence are able to profit a little from drill. Their weakness comes to light in their inability to perform higher types of mental activity.

The following are examples of the under-estimation of intelligence and school ability of children who are under-age for their grade:

M. L. Girl, age 11–2, mental age "average adult" (16), sixth grade, school work "superior," teacher's estimate of intelligence "average." Teacher credits superior school work to "unusual home advantages." Father is a college professor. The teacher considers the child accelerated in school. In reality, she ought to be in the second year of the high school, instead of in the sixth grade.

H. A. Boy, age 11, mental age 14, sixth grade, school work "average," teacher's estimate of intelligence "average." According to the supplementary information the boy is "wonderfully attentive," "studious," and possessed of "all-round ability." The estimate of "average intelligence" is probably due to the fact that he was compared with classmates who averaged about a year older.

K. R. Girl, age 6–1, mental age 8–5, second grade, school work "average," teacher's estimate of intelligence "superior," social environment "average." Is it not evident that a child from ordinary social

environment who does work of average quality in the second grade when barely 6 years of age, and who has an intelligence quotient of about 140, should be judged "very superior" rather than merely "superior" in intelligence?

S. A. Boy, age 8–10, mental age 10–9, fourth grade, school work "average," teacher's estimate of intelligence "average." Teacher attributed school acceleration to "studiousness" and "delight in school work." Our own guess would be that these traits are, rather, indications of unusually superior intelligence.

In a special study of a group of superior children, tested separately from the present investigation, we have found even more striking examples of the difficulty teachers have in recognizing superior ability. One case was that of a boy aged 10–6 with an intelligence quotient of 148. He was in the sixth grade, doing "superior" work there, and yet was judged by the teacher to have "no unusual ability." It was learned from the parents that the boy is distantly related to Meyerbeer, the composer, and that at an age when most children are reading fairy stories, he has a passion for difficult medical literature and text books in physical science.

The question has suggested itself, whether teachers' estimates of intelligence vary in reliability with children of different ages. We have divided our children into three groups, according to age, and have computed for these groups separately the correlation between the intelligence quotient and the teachers' estimates. Ages 5, 6, 7 and 8 were placed in one group; Ages 9, 10 and 11 in another; and Ages 12, 13, 14, and 15 in a third. The coefficients were, in order, .48, .60 and .46. It appears, therefore, that teachers probably make fewer errors with pupils of the middle group, though the difference is not great.

Such facts as we have set forth in this chapter suggest that, while the judgment of a teacher regarding a

child's school success and intelligence may, if properly safeguarded, afford valuable data to supplement the results of the intelligence test, the assistance is more likely to be in the other direction; more often it is the test which can keep us from being misled by the erroneous judgment of the teacher.

SUMMARY

1. The correlation between intelligence quotient and the quality of the school work as judged by the teachers is .45. An examination of the marked cases of disagreement between intelligence quotient and school work shows that these are due largely to the teachers' neglect of age differences in estimating quality of school work.

2. The correlation between intelligence quotient and the teachers' rankings according to intelligence is .48. Detailed study of the cases of disagreement justifies the conclusion that they are due mainly to certain constant errors to which teachers are subject in estimating a child's intelligence. Here, as in judging quality of school work, the most common error is that of overlooking age differences. Teachers judge intelligence mostly by the quality of school work in the grade where the child happens to be located. This results in over-estimating the intelligence of older, retarded children, and under-estimating the intelligence of the younger, advanced children.

3. The wider disagreements between intelligence quotient and grade status are confined chiefly to those children who are superior to, or below the average in ability. The explanation for this has been found in the fact that the tendency of the school is to promote children by age rather than by ability. Those who

have an intelligence quotient between 96 and 105 are hardly ever more than one grade removed from the location which is normal to their mental age.

4. The child with an intelligence quotient of 120 or above is rarely found below the grade for his chronological age, and occasionally he is one or two grades above. Wherever located, his work is nearly always superior, and the evidence suggests strongly that this superiority of school work would continue even if extra promotions were granted. Superior children are seldom allowed to reap the advantage, in school progress, to which their superiority fairly entitles them.

5. The child of 70–79 intelligence quotient never does satisfactory work in the grade where he belongs by chronological age. After the age of 8 or 9 years such a child is usually found doing "very inferior" to "average" work in a grade two to four years below his age.

6. Although the disagreements between intelligence quotient on the one hand and grade progress, quality of school work, and teachers' estimates of intelligence on the other hand would at first seem to justify serious misgivings as to the value of the intelligence scale, these same disagreements, on closer examination, are found to offer the strongest evidence in support of the test method.

CHAPTER VII

THE VALIDITY OF THE INDIVIDUAL TESTS

Criteria of a Test's Validity

The first task in the construction of an intelligence scale is to select tests which are really tests of intelligence, tests which are not too much influenced by age, home environment or school instruction apart from native endowment. There are three criteria which a test must satisfy before it can be accepted as a valid measure of intelligence. In the first place, since we know that intelligence is to a certain extent a function of age, a test to be valid must show an increase from year to year in the percentage of unselected children that pass it. This is the criterion on which Binet chiefly relied. Nearly all the tests which he finally included in his scale satisfy this criterion fairly well, though some show a more rapid increase than others.

This, however, is not sufficient. Many other traits besides intelligence are also functions of age. Height, weight, chest girth, length of forearm, in fact, any physical trait influenced by growth would show a steady increase from age to age in the percentage that pass a given standard. Yet it is easy to show that tests of this kind have no place in an intelligence scale. If 100 unselected 10-year-olds were measured for length of forearm it would of course be found that the average for all considerably exceeds that for 9-year olds; but it would also doubtless be found that this is little if any more true of 10-year-olds who have superior intelligence than of 10-year-olds who have inferior inteligence. That is, although intelligence and

length of forearm are both functions of age, they have no direct relationship to each other. Such a test would not be found coherent with any already existent intelligence scale. Similarly, if a given test in the Binet series does not agree with the scale as a whole, if 10-year children who by the scale have 11-year intelligence do not pass it any more frequently than those 10-year children who have 9-year intelligence, then either this test is worthless or the scale as a whole lacks validity. The entire scale must be coherent.

But coherency and age-increase in the percentage that pass do not themselves guarantee the validity of a series of tests for the measurement of intelligence. A set of tests made up of a great variety of physical measurements might very well satisfy both of these criteria. If we have no already existing intelligence scale with which to compare an individual test, then we must compare the test with intelligence as otherwise estimated, for example, with teachers' rankings. If children who are ranked as intelligent succeed with it no better than those who are ranked as dull, then the test is of doubtful validity.

For our present purposes the third criterion may be left out of account in judging the validity of individual tests of the Binet scale. It has been amply demonstrated that the scale as a whole gives a fairly reliable index of intelligence. Its results always show a reasonably high correlation with intelligence as judged by teachers or other observers. We have already shown that its correlation with school success is fairly high, particularly when allowance is made for certain tendencies to error in the estimation of school success. Its use with feeble-minded children in institutions has been especially convincing. Long-continued observa-

tion of such children rarely necessitates any serious correction of its verdict.

As for the first essential—the requirement of an increase in the percentage of passes from year to year—it is evident from all the available statistics that all the tests which we have included in the revision meet this criterion in a fairly satisfactory way. Some tests show more rapid increases than others, but not one is passed by equal percentages in three successive years.

Accordingly, the criterion of most importance for our purpose is the second one—that of coherency. Since we know that the scale as a whole is fairly reliable, we can measure each individual test against the entire scale. A test which gives results out of harmony with the results of the scale as a whole can not be considered a satisfactory test of intelligence, whatever increase it may show in the percentage that pass it from year to year. This increase might be due to other factors than intelligence, such as school instruction or the incidental experiences which come with increasing age.

One way of applying this criterion would be to classify all our subjects by mental age as determined by the scale and then note the number that pass a given test at successive mental ages. This method gives valuable information, and we have had to rely on it to a large extent in evaluating and placing the tests of the upper-year groups. It has, however, one objectionable feature; the results are more or less influenced by age, apart from intelligence. Children of 8-year mental age, for example, range in chronological age all the way from 5½ or 6 years to 11 or 12 years, and it is conceivable that these large age differences might have a considerable influence on the number that pass a given

test. The only way to separate the influence of intelligence from that of age is to take a large number of unselected children of one chronological age and find the percentage of passes separately for the bright and dull children of that group.

Correlation of the Individual Tests with Intelligence Quotient

Following the foregoing plan, we have divided the children of each chronological age into three groups according to magnitude of intelligence quotient. We have placed in the middle or normal group those children of a given age having an intelligence quotient between 96 and 105, in the inferior group those with an intelligence quotient below 96, and in the superior group those with an intelligence quotient of 106 or above. At most of the ages this gives three groups of about the same size. Had we tested a larger number of children of each age, it would probably have been better to place more children in the middle group, say all between 91 and 110 intelligence quotient. This would have heightened the contrast between the inferior and superior groups. However, such a plan would have placed about 60 percent of our cases of a given age in the middle group and left only about 20 percent, only 12 to 20 children, for each of the other groups. Accordingly, in order to obtain three groups of nearly equal size we have included in the middle group those with an intelligence quotient between 96 and 105.

Table 39 shows the results of this comparison. The figures are in all cases the percentages of passes for children of the chronological age in which the test is located. The three columns give these percentages for children with intelligence quotient 95 or below, 96–105, and 106 or above, respectively.

TABLE 39

	Percent passing below I. Q. 96	Percent passing I. Q. 96–105	Percent passing above I. Q. 105
V			
1. Weights	53	75	84
2. Colors	58	75	89
3. Aesth. Comp.	58	69	84
4. Def. by use	52	69	79
5. Patience	47	75	95
6. Commissions	43	75	90
Al. Age	53	69	95
VI			
1. Right and left	43	70	89
2. Mutilated pict.	32	66	86
3. Thirteen pennies	40	77	96
4. Comprehension (2)	50	68	85
5. Four coins	54	79	83
6. 16–18 syllables	57	65	78
Al. A.M. and P.M.	64	82	86
VII			
1. Fingers	53	68	85
2. Descript. of picture	48	52	80
3. Five Digits	62	74	80
4. Bow knot	43	71	88
5. Differences	48	74	95
6. Diamond	38	58	82
Al. Days of week	33	62	85
Al. Three Digits Reversed	39	55	75
VIII			
1. Ball and field	48	60	69
2. 20–1	35	55	83
3. Comprehension (3)	52	80	80
4. Similarities	44	57	83
5. Superior def.	44	60	80
6. Vocabulary (20)	26	57	74
Al. Six Coins	53	64	71
Al. Dictation	65	90	100
IX			
1. Date	48	68	88
2. Weights	37	58	82
3. Change	39	60	73
4. Four Digits Reversed	45	62	88
5. Three words	49	67	91
6. Rhymes	39	67	88
Al. Months	39	55	85
Al. Stamps	40	76	91
X			
1. Vocabulary (30)	20	60	80
2. Absurdities	35	60	84
3. Designs	45	57	70
4. Reading and rept.	45	63	84
5. Comprehension (4)	25	64	76
6. Sixty Words	30	70	76
Al. Six Digits	55	67	81
Al. 20–22 Syllables	45	57	79
Al. Healy's form board	55	67	76
XII			
1. Vocabulary (40)	36	70	76
2. Abstract words	40	60	79
3. Ball and field	36	65	83
4. Disarranged sentence	45	60	86
5. Fables (score 4)	39	70	86
6. Five Digits Reversed	48	75	79
7. Picture interpreted	21	60	75
8. Similarities	47	65	79
XIV			
1. Vocabulary (50)	42	76	85
2. Induction	41	59	71
3. President and K.	22	62	86
4. Prob. of fact	44	55	76
5. Arith. reas.	25	55	81
6. Clock*	59	83	91
Al. Seven Digits	50	62	67

* Figures for clock problem are on basis of 1 success out of 2 trials.

Table 39 contains not a few surprises. Some of the tests which have been most criticized for their supposed dependence on extraneous factors here show a high correlation with intelligence. Among these are "days of week," "stamps," "13 pennies," "president and king," "rhymes," "age," "right and left" and "picture interpretation." Others having a high reliability are "vocabulary," "arithmetical reasoning," "giving differences," "diamond," "date," "reversed digits," "fables," "disarranged sentences," "60 words," "mutilated pictures" and "absurdities."

Among the poorest are "repeating digits" (direct order), "naming coins," "morning and afternoon," "definition by use," "designs" and "aesthetic comparison." Three of the tests which had been included in our original trial series correlated so little with intelligence that we have thrown them out. They are Binet's suggestion test (in his 1911 revision), Healy and Fernald's Construction Puzzle B, and our test of drawing an apple with a pencil through it. The figures for these tests are as follows:

TABLE 40

	Percent passing below I Q 96	Percent passing I Q 96–105	Percent passing above I Q 105
Binet's suggestion test (Yr. X)	80	62	89
Healy and Fernald Construction Puzzle B (Yr. XII)	75	75	65
Drawing apple and pencil (Yr. VIII)	62	63	57

Tests that correlate moderately with intelligence include "arranging weights," "three commissions," "colors," "fingers," "picture description," "months," "change," "similarities," "superior definitions," "read-

ing for memories," "abstract words," "induction test," "problems of fact," "clock" and "bow-knot."

A test that makes a good showing by this method of comparison, whatever it appears to be from mere inspection, is a real measure of intelligence. Henceforth, it stands or falls with the scale as a whole. That so few of the tests should fail to show a reasonably high correlation with intelligence is striking evidence of the ingenuity and psychological insight of Binet.

It is especially interesting to compare the facts set forth in Table 39 with Meumann's classification of the tests as "tests of maturity," "tests of *milieu*," and "tests of endowment." Of the good or superior tests, Meumann has classified the following as tests of *milieu:* "counting pennies," "colors," "vocabulary," "days of week," "months," "date," "right and left," "20–0," "ball and field," and "60 words." Other excellent tests of intelligence he classifies as tests of maturity. In the light of our results it is hardly necessary to enter into the *a priori* criticisms which Decroly and Degand, Ayres, and others have made of certain of the tests. The classification and criticism of tests by mere inspection may form an interesting pastime, but it can hardly be taken seriously as a contribution to science.

It is not implied that a test which makes a good showing in Table 39 is entirely free from other influences than intelligence. On the contrary, age and environment may affect almost every test to a greater or lesser degree. To determine the exact extent to which this may be true for even a single test, would require an extensive investigation.

The Influence of School Instruction

We have the following data to offer regarding the influence of school instruction upon certain of the tests in the highest groups. A comparison was made of the percentages of passes made on individual tests by "educated" and by "little-educated" adults who tested at the "average adult" level, that is, between 15 and 17 mental age. Of the adults tested by Knollin, Johnson, Zeidler and Terman there were 28 of the "average adult" level who had progressed through the high school or beyond and 33 who had not gone beyond the eighth grade. Of those who reached the "superior adult" level (17–19 mental age) there were 17 of high-school education and 15 of common-school education or less. Table 41 shows the percentages of passes of these two groups on each test in the four highest groups of tests.

The striking thing in this table is the evident lack of any significant influence of school training on the ability to pass most of the tests. The only ones in which the high-school group shows a marked superiority are "sense of selection," "ingenuity," "abstract pairs" (XVI) and "problems of fact." The common-school group has the advantage in "ball and field," "5 digits reversed," "6 digits reversed," "7 digits" (direct order) and "26–28 syllables." "Vocabulary," "fables," "paper cutting" Binet, "clock," "arithmetical reasoning," and "disarranged sentences" show no significant difference between the two groups.

We have often been warned by skeptical friends that size of vocabulary is determined by schooling and not by intelligence, that success with the fables depends wholly upon moral instruction, that "arith-

TABLE 41

PERCENTAGES OF AVERAGE AND OF SUPERIOR ADULTS WHO PASS THE SEVERAL TESTS OF AGE XII AND ABOVE

Test	"Average Adults" (15–17 mental age)		"Superior Adults" (17–19 mental age)	
	H. S. Group	Com. Sch. Group	H. S. Group	Com. Sch. Group
XII				
1. Vocabulary (40)	100	100	100	100
2. Abstract words	94	97	100	100
3. Ball and Field	70	85	84	90
4. Disar. Sentences	94	91	100	100
5. Fables (score 4)	90	94	100	95
6. Five digits reversed	74	92	93	95
7. Pict. interpret	97	84	100	100
8. Similarities	97	85	100	90
XIV				
1. Vocabulary (50)	100	97	100	100
2. Induction	97	85	100	95
3. Pres. and king	89	83	100	87
4. Prob. of fact	97	85	100	100
5. Arith. reas	89	79	94	96
6. Clock*	91	85	100	95
Al. Seven digits	60	90	93	87
XVI				
1. Vocabulary (65)	90	85	100	100
2. Fables (8)	79	74	86	82
3. Abstract pairs	91	73	100	100
4. Box problem	51	76	86	95
5. Six digits reversed	34	61	73	77
6. Code	38	26	75	67
XVIII				
1. Vocabulary (75)	35	30	100	83
2. Paper cutting	30	28	77	89
3. Eight digits	30	34	73	67
4. Sense of selection	63	28	81	57
5. Seven digits reversed	no data	no data	no data	no data
6. Ingenuity	42	20	85	70

* Figures for clock test are on the basis of 1 success in 2 trials.

metical reasoning" and the "paper-cutting" test will be greatly influenced by recency of instruction in arithmetic and geometry, respectively, and that the ability to repeat sentences and digits depends almost purely upon that rote memory which is so diligently cultivated in the schools. From the data presented above, however, it is evident that criticisms based on off-hand opinion have little value.

The Influence of Age and Experience

The influence of maturity, apart from native endowment, is difficult to isolate. We can, of course, compare older and younger individuals of the same mental age with respect to the percentages that pass individual tests, but this gives us the combined influence of maturity and experience. The 16-year-old of 10-year intelligence is not only more mature than the 10-year-old of 10-year intelligence; he has had also the advantage of 6 years additional experience and opportunity to learn. The best we can do, however, is to treat these two influences as one, which we may call the "age factor," meaning thereby the combined influence of age, as such, and of the experience which goes with age.

The influence of the "age factor" may be seen by comparing Williams' delinquents and Knollin's unemployed of a given mental age with unselected school children of that chronological age. Such comparisons were made for the Mental Ages 9 to "average adult." It will be remembered that most of the delinquents were between 12 and 21 years of age and that most of them were mentally retarded (nearly a third were classed as feeble-minded). The unemployed were somewhat less inferior and ranged from 20 to 65 years

of age, with a median of 34 years. Accordingly, if the delinquents and unemployed of the mental age 10 years differ greatly from unselected 10-year children, the difference may be attributed to the superior chronological age of the former.

A good deal of such influence was found in certain of the tests. The influence, however, is not always in the same direction. With some tests, age tends to increase, with others to decrease, the percentage of successes. Indeed, there are more marked cases of negative influence than of positive. Among those negatively influenced are: "rhymes," "six digits," "60 words," "20–22 syllables," "disarranged sentences," "sense of selection" and "fables"; also, though to a less extent, "date," "three words," "designs," "reading for memories," "26–28 syllables," "five digits reversed" and "similarities" (three things). In finding rhymes, for example, Williams' older delinquents and Knollin's unemployed adults do no better at Mental Age 13–14 than unselected children of 10 years. Almost as great a difference in the same direction is found in the test of repeating six and seven digits, naming 60 words, repeating five digits backwards, and giving sense of selection.

The positive influence of age, *i. e.*, that which causes increase in the percentage of passes, is in evidence with the following tests: "induction," "physical relations," "months," "vocabulary," "comprehension" (Age X), "making change," "problems of fact" and "enclosed boxes." In "comprehension" (Age X), for example, Knollin's adults of 10-year mental level do about as well as unselected children of 11 years. In the induction test the adults of 13-year mental level are far superior to unselected children of 14 years, and the

difference is almost as marked for most of the other tests listed above.

Kuhlmann's data on the influence of age are in some respects the most valuable yet offered on this subject.[1] His method was to compare older and younger feeble-minded children with reference to the percentage that passed each individual test in the Binet 1908 scale from Year III to Year X. Evidence of more or less influence of age was found with the following tests: "pointing to eyes, nose and mouth," "giving last name," "repeating three digits," "copying a square," "counting four pennies," "definitions by use"; "repeating five digits," "naming four coins," "reading for two memories," "naming days of the week" and "naming the months." In only three of the tests, however, was the influence great enough to amount to a displacement of the test by as much as one year. These were "repeating 3 digits," "definitions by use" and "naming four coins."

Pintner and Paterson made a similar comparison of feeble-minded children above and below the chronological age of 15 years with respect to the percentage at each mental age that passed two tests—"days of the week" and "months of the year." Their results, which are given in Table 42, show a fairly large influence of age. The mental ages were determined by the Goddard revision.[2]

---

[1] Kuhlmann: The Binet-Simon tests of intelligence in grading feeble-minded children. *J. of Psycho. Asthenics*, 16: June, 1912, pp. 173–193. See especially pp. 182–185.

[2] R. Pintner and D. G. Paterson: Experience and the Binet-Simo n tests, *The Psychological Clinic*, 8: Dec., 1914, pp. 197–200.

TABLE 42

NUMBER OF FEEBLE-MINDED ABOVE AND BELOW 15 YEARS OF AGE THAT PASSED CERTAIN TESTS. (FROM PINTNER AND PATERSON)

Tests	Chron. Age	Mental Age											
		4		5		6		7		8		9	
		No.	Per-cent	No.	Per-cent	No.	Per-cent	No.	Per-cent	No.	Per-cent	No.	Per-cent
Days of Week	Below 15	21	4.8	52	27	95	59	100	87	72	98.6	31	100
	Above 15	28	14.3	50	54	77	80.5	127	98.5	128	100	101	100
Month of Year	Below 15			19	5.3	66	10.6	100	35	71	77.5	31	87
	Above 15			36	13.9	70	28.6	112	68	128	92.3	101	93

The following tests are also thought by Pintner and Paterson to be subject to age influence, although no data are offered in support of their belief: "counting 13 pennies," "copying a square," "diamond," "naming colors," "telling forenoon from afternoon" and defining in terms of use." On the other hand, they present data which show that with the Knox Cube Test and Knox Feature Profile Test, younger feeble-minded children do better than older feeble-minded children of the same mental age.[3]

Chotzen, who made a similar comparison between the older and the younger feeble-minded children of each mental age, found a marked influence of age only in the following tests: "writing from copy, "writing from dictation," "reading for two memories" and naming the days of the week." A slight influence of age was found in the case of the following; three com-

[3] Pintner and Paterson: The factor of experience in intelligence testing, *The Psych. Clinic*, 9: 1915, pp. 44–50.

mands," "counting backwards, and repeating 16 syllables." It will be noted that those that showed marked influence of age are all tests which relate largely to matters of information, particularly school information. According to Chotzen's data, age, apart from intelligence, plays no part in the case of tests "that demand ability to judge and to combine, or with such as make severe demands upon comprehension, such as comparison, problem-questions, noting omissions, or repeating five digits."[4]

Wallin compared epileptics above and below 21 years of age with respect to the following tests: time of naming four colors, time of reading B.–S. selection, memories from reading B.–S. selection, time of naming 60 words, and time for solving the Goddard form board.[5] No important difference was found except in the case of one or two tests. Epileptics above 21 were superior with respect to time required for reading the selection, but those below 21 retained more memories. In the form-board test, the younger subjects were a little superior to the older. Wallin concludes that common age standards can be used, without much error, for both subnormal children and subnormal adults.

It is hard to summarize the foregoing studies, because their data are not wholly comparable. The subjects of Wallin were epileptics and were grouped for comparison into those below and those above 21 years of age. The subjects of Kuhlmann, Chotzen, and Pintner and Paterson were all feeble-minded, but

[4] F. Chotzen: Die Intelligenzprüfungmethode von B.-S. bei schwachsinnigen Kindern., *Zeit. f. Ang. Psych.*, 6: 1912, pp. 411–494. See especially p. 453.

[5] J. E. W. Wallin: The Binet-Simon tests in relation to the factors of experience and maturity, *The Psychological Clinic*, 8: Feb., 1915, 266–271.

were grouped differently by age. Pintner and Paterson took Age 15 as the dividing line, while Kuhlmann made four groups, as follows: Ages 6–10, 11–15, 16–25, 16 and over. Chotzen's figures are of limited value because of the small age differences among his 236 subjects, most of whom were between 8 and 10 years old. Our own comparison is between normal individuals and backward groups of the same mental age. Moreover, only a part of the subjects of our backward groups (delinquents and unemployed) were actually retarded to any serious degree. Our data on the influence of age are also limited to the tests above the middle part of the scale.

Conclusions

1. The influence of the age factor on ability to pass the tests of the Binet scale does not appear to be great enough or frequent enough to affect seriously the accuracy of the scale as a whole.

2. Tests which are significantly easier for older than for younger subjects of a given mental age include the following: "days of week," "months of the year," "reading a selection," "writing from dictation," "naming coins," "making change," "defining in terms of use," "induction," "vocabulary," "physical relations," "problems of fact" and the "hard comprehension" questions.

3. Tests which are significantly harder for older than for younger individuals of a given mental age include: "rhymes," "six digits," "60 words," "20–22 syllables," "disarranged sentence," "sense of selection" and "fables."

Kuhlmann's assumption that age never exerts a negative influence is certainly not valid when "age" is taken in the sense in which it is here employed.

However, the assumption might hold for age in the sense of *maturity apart from experience.* The fact that adults of a given mental age are inferior in certain tests to unselected school children of the same mental age is probably accounted for by the more recent practice of school children in the type of operation called for by those tests. This explanation would seem to hold especially for such tests as "memory for digits," "memory for sentences," "three words," "rhymes," "reading for memories," "sense of selection," "fables," etc.

4. With regard to memory, the data are contradictory. Kuhlmann finds a positive influence of age for "three digits" and "five digits"; Chotzen classes "memory for digits" with the tests uninfluenced by age, while with our own subjects age has a markedly *negative* influence on memory for six and for seven digits. In the test of "reading for memories" our older subjects of a given mental age read more rapidly than the younger but give fewer and less literal memories. The adult unemployed were especially inferior to school children of the same mental age in giving literal memories—a characteristic which was strikingly evident also in the test of giving the sense of a selection heard. In the latter test the almost irresistible tendency of the older subjects was to give a critical reaction or comment, rather than an epitome of the thought contained in the selection.

5. The data available throw little light on the influence of maturity apart from the influence of experience which comes with increased age. It is not improbable that the influences we mention are due less to maturity, as such, than to the experience incident to age.

6. With regard to the validity of the tests in general, the two most important criteria are: (*a*) increase in the percentage of unselected children that pass from year to year, and (*b*) coherency. As for the former, our data show that all the tests included in our revision satisfy this criterion very well.

7. In order to test the coherency of the scale, each test from V to XIV was measured against the scale as a whole. The results show that all the tests, including those which have been most criticised, are reasonably good tests of intelligence, although not all are of the same value.

CHAPTER VIII

CONSIDERATIONS RELATING TO THE FORMATION OF AN INTELLIGENCE SCALE

The Selection of New Tests

In devising tests of intelligence it is necessary to be guided by certain assumptions regarding the nature of intelligence. To adopt any other course is to depend for success upon happy chance. However, it is impossible to arrive at a final definition of intelligence on the basis of *a priori* considerations alone. To demand, as critics of the Binet method have sometimes done, that one who would measure intelligence should first ascertain exactly what intelligence is, is quite unreasonable. As Stern points out, electrical currents were measured long before their nature was well understood. Similar illustrations could be drawn from the processes involved in chemistry, physiology and other sciences. In the case of intelligence it may be truthfully said that no adequate definition can possibly be framed which is not based primarily on the symptoms empirically brought to light by the test method. The best that can be done in advance of such data is to make tentative assumptions as to the possible nature of intelligence and then subject these assumptions to tests which will show their correctness or incorrectness. New hypotheses can then be framed for further trial, and thus gradually we shall be led to a conception of intelligence which will be meaningful and in harmony with all the ascertainable facts.

This, in fact, was the method of Binet. Only those unacquainted with Binet's more than fifteen years of labor preceding the publication of his intelligence scale would think of accusing him of making no effort to analyze the mental processes which his tests are designed to measure. It is true that many of Binet's earlier assumptions proved untenable, but he was then always ready to acknowledge his error and to plan a new line of attack. The exceptional candor and intellectual plasticity which he displayed in his patient researches with intelligence tests should serve as a splendid example to all who would contribute to this difficult field of psychology.

It is not our purpose to enter here into a criticism of the various tentative definitions of intelligence. Among the most noteworthy of such definitions, or conceptions, are those of Ebbinghaus, Binet, Meumann, Stern, and Spearman. These conceptions may be briefly characterized, in order, as follows:

Binet's conception of intelligence emphasizes three characteristics of the thought process: (1) its tendency to take and maintain a definite direction, (2) the capacity to make adaptations for the purpose of attaining a desired end, and (3) the power of autocriticism.[1]

According to Ebbinghaus, the essence of intelligence lies in comprehending together, in a unitary, meaningful whole, impressions and associations which are more or less independent, heterogeneous or even partly contradictory. "Intellectual ability consists in the elabor-

[1] See Binet and Simon: L'intelligence des imbeciles. *L'Annee psychologique*, 15, 1909, pp. 1–147. The last division of this article is devoted to a discussion of the essential nature of the higher thought processes and is a wonderful example of that keen psychological analysis in which Binet was so gifted. This and other conceptions of intelligence are reserved for treatment elsewhere.

ation of a whole into its worth and meaning by means of many-sided combination, correction, and completion of numerous kindred associations.—It is an *activity of combination.*"

Meumann offers a two-fold definition. From the psychological point of view, intelligence is the power of independent and creative elaboration of new products out of the material given by memory and the senses. From the teleological point of view, it involves the ability to avoid errors, to surmount difficulties, and to adjust to environment.

Stern defines intelligence purely in teleological terms as "the general capacity of an individual consciously to adjust his thinking to new requirements: it is general adaptability to new problems and conditions of life."

Spearman, Hart and others of the English school define intelligence as a "common central factor" which participates in all the special mental activities. This factor is explained in terms of a psycho-physiological hypothesis of "cortex energy," "cerebral plasticity," etc.

These conceptions are only to a slight extent contradictory or inharmonious. They differ mainly in the point of view from which they are framed and in the location of emphasis. Doubtless each expresses a part of the truth and no one all of it. In devising new tests it is well to keep in mind all of these conceptions and others as well. We may thus be spared much fruitless testing of mental processes which are little concerned in intelligence. All the conceptions given agree in locating intelligence chiefly among the higher and more complex processes, as distinguished from the elementary and simple. Tests of sensory discrimina-

tion, reaction, time, etc., have proved themselves all but valueless in the measurement of intelligence. Seashore's suggestion that in order to avoid influences of training we should seek for intelligence tests among processes which mature so early as not to show age differences after the early years of childhood, finds no warrant in any of the published results of intelligence testing.

When a test is found which from inspection appears capable of bringing to light individual differences in the mental traits involved in any reasonable conception of intelligence, it should be tried. As we have already indicated, the best way to do this is to apply the test to a large number of unselected children of different ages. If the percentage of success increases but little or not at all with age, the test may as well be discarded. On the other hand, if the proportion of successes increases significantly with age, then the test is a valid measure of intelligence *unless the increase is due to maturity or experience apart from intelligence.* To ascertain whether intelligence, rather than maturity or experience, is responsible for the increase it is necessary to compare the intelligence of those children of a given age who pass the test with the intelligence of children of the same age who fail it. This comparison is readily made if an intelligence scale is available which is known to have a certain degree of validity. It is the method we have employed in the preceding chapter. When no such scale is available for comparative purposes we can still check up the validity of a test by utilizing school records, teachers' estimates of the children's intelligence, and other criteria. It is perhaps always best to use those criteria in addition to measuring the individual test against an existing scale.

It can not be too strongly emphasized, however, that a test may under ideal laboratory conditions be a valid measure of intelligence and yet not be usable under the limitations of time and equipment which usually prevail in Binet testing. To be widely serviceable a test should demand only the simplest material or apparatus, should require at most but a few minutes of time, and should lend itself well to uniformity of procedure in application and scoring.

The Assembling of Tests into a System

The Binet scale has often been criticised as a heterogeneous aggregation of tests without plan or system. This criticism is based on the assumption that the only way to measure intelligence is first to measure separately the individual mental functions involved in intellectual processes and then to summate the results. It would thus be necessary to have separate scales for such functions as accuracy of perception, memory, logical association, reasoning, judgment, etc., or possibly several scales for each one of these functions, in order to measure its efficiency with different kinds of material.

From this point of view, the Binet scale is indeed a "motley array" of tests. Although there are several memory tests, no effort is made thoroughly to test any kind of memory. There are three tests that involve drawing with pen or pencil, but there is no pretense of testing thoroughly any particular functions involved in drawing from copy or memory. The same may be said of various tests of association, mastery of language, comprehension, reasoning, etc. How, it is often asked, can the scale measure intelligence as a whole when it offers no reliable measure of any single aspect of intelligence?

This is the point of view of "faculty psychology," which, far from being defunct, has really given direction to much of the current work in intelligence testing. It was the point of view which long controlled the work of Binet, who, like others, began by attempting to get at intelligence by measuring memory, attention, sense discrimination and other individual functions. It was only after years of exploration along the old lines that he finally broke away from them and undertook, so to speak, to triangulate the height of the tower without first getting the dimensions of the individual stones which made it up. The assumption that it is easier to measure a part, or one aspect, of intelligence than all of it, is fallacious in that the parts are not separate parts and can not be separated by any refinement of experiment. They are interwoven and intertwined. Each ramifies everywhere and appears in all other functions. The analogy of the stones of the tower does not really apply. Memory, for example, cannot be tested separately from attention, or sense discrimination separately from the associative processes. After vainly trying to disentangle the various intellective functions Binet decided to test their combined functional capacity without any pretense of measuring the exact contribution of each to the total product. Intelligence tests have been successful just to the extent to which they have been guided by this aim.

Memory, attention, imagination, etc., are terms of structural psychology. Binet's psychology is dynamic. He conceives intelligence as the sum total of those thought processes which consist in adaptation. This adaptation is not explicable in terms of the old mental faculties. No one of these "faculties" can explain a

single thought process, for such process always involves the participation of many "faculties," whose separate rôles are impossible to distinguish accurately. Instead of measuring the intensity of various mental states (psycho-physics), it is more enlightening to measure their combined effect on adaptation. Using a biological comparison, Binet says the old "faculties" correspond to the separate tissues of an animal or plant, while his own "scheme of thought" corresponds to the functioning organ itself. Binet's psychology was a functional rather than a structural psychology.[2]

Binet's conception of "general intelligence," although expressed in psychological terms, harmonizes well with Spearman and Hart's psycho-physiological conception of intelligence as depending upon general cerebral efficiency, or "cortex energy." The assumption is common to both that a common factor, "intelligence" (Binet) or "general ability" (Spearman), enters into every specific mental performance of an individual, and that by testing a subject with a large variety of mental tasks the special factors involved are canceled and the general factor revealed. If Binet's theory of intelligence is valid, we must reject those criticisms which condemn the scale for its failure to test out and follow up individual mental functions from childhood to maturity. The incorporation in the scale of heterogeneous tests and the partial lack of consecutiveness in their arrangement were the result, not of accident, but of a well-defined theory. Whether the underlying theory is correct, is a matter to be determined by patient research.

[2] See Binet and Simon: L'intelligence des imbeciles, *L'Annee Psychologique*, 15, 1909. Especially pp. 143–147.

It is unfortunate that off-hand criticisms of the Binet method have been so confidently voiced by psychologists who have evidently not taken the trouble to acquaint themselves with Binet's work. One of these criticisms is to the effect that since the tests of successive years in the scale do not follow up the same functions, the different "mental ages" are therefore not comparable; that we do not know, for example, how a child of Mental Age VIII differs from one of Mental Age VI, since the two mental ages have not been earned by tests of the same functions. The criticism implies that "mental age," in terms of the Binet scale, is meaningless. We would urge, however, that whether the different mental ages have meaning and are comparable is a question which can only be settled experimentally. We have already shown (Chapters III and VI) that the evidence strongly supports the validity of the Binet method. Knowing that a given child of 6 years has a mental age of 5 years by the Binet scale, we are able to forecast fairly accurately the mental age the child will have at 12 years, and even the probable degree of his school success. The proof of the Binet method is in the fact that it works so well. Investigation alone will determine whether a scale made up on the plan advocated by Rossolimo, Thorndike, Yerkes and others (that is, a scale designed to follow up and measure individual functions separately) is capable of furnishing an index of intelligence any more reliable or any more meaningful for purposes of age comparison than that secured by the Binet scale. Indeed, it is conceivable that a scale whose various year-groups of tests were composed with studied effort to make them dissimilar and non-comparable might still give intelligence

quotients of high validity at the different ages. It is to be hoped that someone will make such an experiment.

Whatever the facts may prove to be, it is probably desirable for the present to include several varieties of tests in each year group, and as far as possible to avoid extreme dissimilarity in the general character of the tests of successive years. It may even be desirable, when it is convenient, to have the same test or the same type of test recur at different age levels. Binet, himself, frequently followed this plan. It is not at all certain, however, that it is necessary for the purpose of measuring general intelligence to go much farther than did Binet in the direction of testing out the special mental functions.

The Location and Scoring of Tests

The percentage of correct responses necessary for locating a test is a much-debated question. Binet's standard was a shifting one, varying from 60 to 90 percent according to the upward trend of the curve for the test in question. Goddard, Kuhlmann, Bobertag, Stern and Meumann adhere strictly to the 75 percent standard. Bobertag and Stern seek to justify this standard by the (supposed) fact that it gives a fairly normal distribution of mental ages and causes approximately 50 percent to test "at age," that is, within 12 months of normal. It is true that the 75 percent standard has this effect at a certain part of the scale. It is evident, however, from the data presented in Chapter III, that the relation between the 75 percent standard and the 50 percent testing "at age" is a purely accidental one. That it should hold at any point is due to the chance ratio that obtains for a

certain period between chronological age and the rate of growth. There can be no correct scale which will cause 50 percent of unselected children to test "at age" at all the age levels. If the scale is accurate it appears that the proportion of "at age" cases among 4-year-olds will be about twice as great as among 8-year-olds and about three times as great as among 12-year-olds. That is, the curve of distribution of mental ages of unselected children becomes more and more flattened as we ascend to the higher age groups.

This last statement is only another way of expressing the familiar fact that, relative to the mental development already attained, a year of growth amounts to less in the upper than in the lower years. A year of mental growth added to the mental age of 5 years amounts to an increase of 20 percent, while a year added to the mental age of 10 years is an increase of only 10 percent. It is evident, therefore, that if a scale contains a group of tests at each age level, these groups become progressively closer together up the scale, until finally, when the point of mental maturity has been attained, the successive age groups are not separated at all. The distance between the 5-year and 6-year tests is as great as that between the 10-year and 12-year tests, and 21-year tests are probably not separated from 20-year tests by any distinguishable distance whatever.

This crowding up of the age levels in the upper years is of greatest significance as regards the percentage of passes necessary for locating tests in a scale like the Binet, for it means that the true standard percentage is not uniform for the different age levels, but variable. This will be evident if the following facts are borne in mind:

1. A child does not earn a given mental age, say 8 years, by his performances on the tests of the 8-year group alone. He earns it by certain successes in the 8-year group, plus certain successes in the succeeding year-groups, minus certain failures in the preceding year-groups.

2. Since the year-groups above the given mental age are relatively closer together than the year-groups below that mental age, it is evident that *if the tests of all the different age-groups are located according to the same standard* (say 75 percent of passes), then unselected children of any given age will, on the average, attain more successes above their age than they will suffer failures below their age.

3. In testing unselected children of the various ages with a scale of this kind (that is, a scale having an equal number of tests at each age and all tests located according to a uniform standard) *the ratio between* (a) *the successes above a given age attained by the children of that age and* (b) *the failures below that age suffered by the same children can not be exactly the same at any two age-levels.* This would cause a displacement, and an unequal displacement, of median mental age from median chronological for the children of the upper age-levels.

4. It must be especially borne in mind, that the guiding principle in the formation of an accurate scale is that *the median mental age must coincide with median chronological age for unselected children of each age.*

There are three ways of overcoming these difficulties so as to bring it about that median mental age will coincide with median chronological age: (1) the number of tests in the upper age-groups may be progressively reduced, (2) the tests of the upper age-groups may be located on the basis of a smaller and smaller percent passing; or (3) the two methods may be combined.

Binet's 1911 revision meets the situation in part by omitting certain year-groups of tests altogether. The upper tests of both the 1908 and 1911 series are also somewhat harder for the age in which they occur than are the tests at the lower end of the scale, though this was probably unintentional. The difficulty can really not be satisfactorily solved by a progressive reduction in the number of tests, since tests can only be dropped by wholes. In certain cases, to retain a test would give too much advantage in one direction, while to drop it would throw too much weight in the opposite direction.

In the Stanford Revision we have fewer tests per year in the upper part of the scale than in the lower, but we have not attempted to make the decrease perfectly gradual. Instead, whole year-groups have been omitted. This irregularity was then remedied empirically in three ways: (1) the number of tests in Group XII was increased to eight; (2) higher values were assigned to the tests of the "average adult" and "superior adult" groups; and (3) tests were located in the upper groups on the basis of a smaller percent passing. Changes along these lines were continued until an arrangement was found which caused the median mental age of unselected children of any age to coincide approximately with median chronological age. When the tests had been arranged in such a way as to bring this about, it was found that they were passed by the following percents in the locations assigned them.[3]

[3] The percents from IV to XIV, inclusive, are those found for Stanford unselected cases in 1914–1915. As unselected cases were not available for the "average adult" and "superior adult" tests, the figures for those were obtained by taking the arithmetical average of the percents for the high-school pupils, the business men, Williams' delinquents of the "average adult" level and Knollin's unemployed of the same level. The resulting figure presumably shows what "average adult" intelligence can do with the tests of the "superior adult" group.

TABLE 43

Year Group	Average Percent Passing	Year Group	Average Percent Passing
IV	77	IX	62.3
V	71.3	X	64.5
VI	70.8	XII	62.4
VII	68	XIV	55.6
VIII	63.2	"Av. Adult"	59.8
		"Sup. Adult"	37.4

The percents are the averages for the regular tests only; the alternates are omitted. The figure for XIV does not include the clock problem, owing to a change in the method of scoring employed in the revision, and in "superior adult" the test of repeating seven digits backward is not taken ino account for lack of sufficienct data.

It will be noted that from IV to IX there is a general decrease in the average percent of passes for the tests of successive year-groups. This is what we should expect from the facts set forth. Beyond IX, however, the percents vary more or less irregularly, owing to the omission of XI, XIII, XV, etc. The omission of XI must also affect the placing of tests in IX and X. If tests were included at XI, then the tests of IX and X would either have to be made harder or else shifted downward in order to keep the median mental ages at these levels from running too high.

It is evident that the percents beyond VIII have no applicability or meaning apart from the present scale as it is actually constituted. A scale constructed somewhat differently, with more or fewer groups of tests in the upper ranges, or with slightly different values assigned to the tests of the different year-groups, would require other percents for the correct location

of the tests. Similarly, the percents below IX would lose general significance if any of the year-groups were omitted. As stated elsewhere, we have not been guided by theoretical considerations in locating the tests, but by the one purpose of securing an arrangement which would give median mental ages as nearly as possible equal to the median chronological ages. Although we have endeavored to keep the tests of a given age-group as nearly as possible of the same difficulty, it must always be remembered that the correct location of a test does not depend wholly upon the percent that pass it any one year.

It would be much more satisfying if the difficulties of the age-grade method could be disposed of on the basis of purely logical considerations, without the necessity of trial and error. The problem is so involved, however, that such a solution is not easy. For example, to make a correct scale by reducing in exactly the right proportion the number of tests in the upper years, or by introducing a variable standard for locating the tests in the different year-groups, would be impossible on logical considerations unless we had exact knowledge as to the normal rate of growth at every point.

It would doubtless be possible to construct an intelligence scale by a method based purely on logical consideration. Mr. Otis[4] has formulated such a plan which involves, among other features, the location of each test at the age where it is passed by 50 percent of unselected children. However, Mr. Otis' suggestions are hardly applicable to a scale like that of Binet and so need not be set forth here.

[4] In a thesis on "The logical and mathematical aspects of intelligence testing by the Binet method," *Psychological Review*, March and May, 1916.

Point Scales

Yerkes has strongly emphasized the advantages of a point scale as contrasted with the age-grade method of Binet. Three advantages are claimed for the point scale: (1) that it becomes better standardized, the more widely it is used, however much the results secured lack agreement; (2) that the point scale admits of more ready comparison of intelligence norms of the sexes, different races, and the various social classes; (3) that it more conveniently permits the giving of partial credit for partial success in a test.

These claims are less valid than they at first appear. In the first place, the lack of agreement in Binet data is unquestionably due in large measure to lack of uniformity of procedure in giving and scoring the tests, that is, to sources of error which are no more avoidable in the use of a point scale than any other. We have tried to remedy this defect for the Binet scale by preparing a very detailed guide.

In the second place, it can not be admitted that the Binet norms necessarily become more insecure or "muddled" with the accumulation of data collected by satisfactory procedure. Such data can, in fact, be readily utilized for purposes of revising the scale from time to time. Another way would be to let the scale stand, admittedly imperfect as it is, but to make use of all available results for correcting the intelligence quotients which the scale gives at each age. If it should be found, for example, that the present revision gives for an indefinitely large number of unselected 6-year-olds a median intelligence quotient of 103, instead of the desired 100, it would then be merely necessary to use 103 as the norm instead of 100. Similarly with an error in either direction at any other age.

Finally, as shown in Chapter III and as is implied above, the Binet method becomes itself a point scale as soon as we express intelligence status in terms of the intelligence quotient. In terms of its "points" (that is, units of intelligence quotient) we may compare sexes, races or social classes as readily as by any other point scale. Moreover, the "points" of the Binet scale are more nearly equal units than can possibly be the case with a point scale whose various tests have had point values assigned to them in an arbitrary and off-hand manner.[5]

The question whether a scale to be valid must give partial credit for partial success in a test is another aspect of the question whether it is necessary for the tests to follow up the development of individual functions. In addition to what we have already said on this matter it may be pointed out that it is impossible entirely to avoid the "all or none" principle in scoring. A subject who is being tested even by the Yerkes-Bridges scale will in many cases lose "points" because his ability in the function tested falls just a little short of that necessary to earn the credit assigned to the test, or part of a test. There is one real advantage, however, in giving as far as it is convenient to do so, partial credit for partial success; it makes it possible to get along with a somewhat smaller number of tests. By following this method, as we have done with the tests of vocabulary, fables, ball and field, etc., we may make a single test serve as two or three, or even more tests.

[5] See Arthur S. Otis: A criticism of the Yerkes-Bridges point scale, with alternative suggestions, *Journal of Educational Psychology*, March, 1917.

Summary

1. In the selection of intelligence tests it is necessary to be guided by a tentative conception as to the nature of the processes to be measured. It is unreasonable, however, to demand that intelligence testing wait upon a complete definition of intelligence. Such a definition can be arrived at only empirically, by the use of tests which bring to light the symptoms of intelligence.

2. Many of the criticisms of the Binet method are based on an unfortunate lack of acquaintance with Binet's psychological work and of the considerations which shaped his system of tests. Whether the age-grade method is inferior to the "profile" method, or what we might call the "special function" method, can be answered only in the light of future researches.

3. In a system like that of Binet there is no single standard as to the percentage who should pass a test in a given year in order to determine its location in that year. Factors involved are: (1) the relative amount of mental growth from year to year, and (2) the number and weighting of tests at the upper levels.

4. The advantages claimed for the point-scale method are questionable. The most important consideration in framing such a scale, namely, that of equalizing the "points," has been overlooked. We have tried to show that the Binet scale itself, properly standardized, becomes automatically a point scale not easily improved upon.

APPENDIX I

The Data Utilized in the Stanford Revision

All of the 2060 tests were made by trained examiners, with the exception of about a fourth of the Terman-Childs cases. As explained in the text, the Terman, Lyman, Galbreath, Talbert, and Cuneo subjects were within two months of a birthday and were as nearly as possible unselected. The Terman and Childs subjects were drawn from social classes perhaps slightly above the average, though this was not fully realized at the time the tests were made. The Terman, Trost and Waddle cases were tested in a trial of the Terman-Childs Tentative Revision of 1912. Of the 310 in the latter group, 50 were tested by Dr. Charles Waddle, State Normal School, Los Angeles, California, 125 by Miss Helen Trost, a student at Stanford University, and the remaining 135 by Terman. They were probably of about average social status. Effort was made to avoid selection, but the precautions taken to this end were not adequate.

The following tables give the percentage that passed each test at each age. Where statistics are given, it will be understood that all, or practically all, of the subjects of the age in question were given the test. A few tests, owing their improper location in the trial series, were not given to all the subjects to whom they should have been given, and in such cases the percents could not be tabulated for the desired range of ages.

Terman's high-school students were from 17 to 20 years of age, were members of the junior and senior classes, and were all doing school work of about average quality.

Dr. Williams' delinquents were inmates of the Whittier State School, one of the three juvenile reform schools of California. It should be noted that the statistics for Williams' delinquents and Knollin's unemployed are based on the classification of these subjects by mental age.

The Hopwood and Houser vocabulary tests were mass (written) tests of pupils from the fourth to the eighth grades in Alameda and Riverside, California. Approximately 30 pupils were tested at each age indicated in each of these series. The Terman vocabulary data from college students were also collected by mass (written) tests. The students numbered 65 and belonged chiefly to the junior and senior classes.

Since the above data were collected, additional tests have been made of 250 unskilled, semi-skilled, and skilled employees, 160 prisoners, 300 additional juvenile delinquents, 130 high-school students, 80 children in an orphanage, 100 kindergarten children, 200 first-grade children, 250 children in other grades, about 100 superior children, 174 children suspected of being feeble-minded, and 140 special-class children. Retests have also been made of nearly 150 children tested by Terman and Childs. The results of these investigations will be published in another monograph.

DATA UTILIZED IN THE STANFORD REVISION

Source of Data	Number at Each Age														Total
	3	4	5	6	7	8	9	10	11	12	13	14	15	16	
Terman, Lyman, Galbreath, Talbert, Cuneo.	10	17	54	117	92	100	113	87	79	83	98	82	46	14	992
Terman and Childs.		29	83	26	29	43	49	33	44	35	17	6	2		396
Terman, Trost, and Waddle.		5	10	22	39	44	34	55	36	25	22	18	10	7	310
Terman's high-school pupils.	(Ages 16–21. Junior and senior classes.)														32
Knollin's business men.	(Ages 25–65 years. Had not attended the high school.)														30
Knollin's unemployed men.	("Hoboes," ages 21–60.)														150
Williams' juvenile delinquents.	(Ages mostly 14–21.)														150
															2060

Percents That Passed Individual Tests in Different Stanford Investigations

(Figures at top of each column refer to ages)

III[1] Parts of the Body

	3	4	5	6	
Terman, Lyman, Ordahl, etc.	80	86	100	100	(Nose, eyes, mouth, hair, 3 of 4)
Terman and Childs.....		89	96	100	(Nose, eyes, mouth, 3 of 3)
Terman, Trost, Waddle....		100	100	100	(Nose, eyes, mouth, 3 of 3)

III[2] NAMING FAMILIAR OBJECTS

	3	4	5	6	
Terman, Lyman, Ordahl, etc.	90	100	100	100	(Key, penny, knife, watch, pencil, 4 of 4)
Terman and Childs		93	100	100	(Key, penny, knife, 3 of 3)
Terman, Trost, Waddle		100	100	100	(Key, penny, knife, 3 of 3)

III[3] PICTURES, ENUMERATION

	3	4	5	6
Terman, Lyman, Ordahl, etc.	80	94	100	100
Terman and Childs		96	100	100
Terman, Trost, Waddle		100	100	100

III[4] GIVING SEX

	3	4	5	6
Terman, Lyman, Ordahl, etc.	80	94	100	100
Terman and Childs		100	100	100
Terman, Trost, Waddle	100	100	100	100

III[5] GIVING FAMILY NAME

	3	4	5	6
Terman, Lyman, Ordahl, etc.	80	94	100	100
Terman and Childs		95	100	100
Terman, Trost, Waddle		75	100	100

III[6] REPEATING 6–7 SYLLABLES

	3	4	5	6
Terman, Lyman, Ordahl, etc.	70	94	100	100

III ALTERNATIVE. THREE DIGITS

	3	4	5	6
Terman, Lyman. Ordahl, etc.	70	87	100	100
Terman and Childs		97	100	100
Terman, Trost, Waddle	100	75	100	100

IV[1] COMPARISON OF LINES

	3	4	5	6
T., L., O., etc.	60	85	97	99
T. and C.		100	100	100

IV[2] DISCRIMINATIONS OF FORMS

	3	4	5	6
T., L., O., etc.	10	70	83	95

IV[3] COUNTING FOUR PENNIES

	3	4	5	6
T., L., O., etc.	20	77	93	98
T. and C.		91	90	100
T., T., W.		80	100	100

IV[4] Copying Square

	3	4	5	6	
T., L., O., etc.	10	71	76	95	(Pencil, 1 of 3)
T. and C.		86	91	95	(Pencil, 1 trial only)
T., T., W.		60	88	100	(Pencil, 1 trial only)

IV[5] Comprehension, 1st Degree

	3	4	5	6
T., L., O., etc.	40	82	93	96

IV[6] Repeating Four Digits

	3	4	5	6
T., L., O., etc.	40	76	83	91

IV Alternative. 12–13 Syllables

	3	4	5	6	
T., L., O., etc.	30	83	85	91	(12–13 syllables, 1 of 3)
T. and C.		51	83	85	(14 syllables, 1 of 1)
T., T., W.		60	75	91	(12–14 syllables, 1 of 3)

V[1] Comparison of Weights

	4	5	6	7	
T., L., O., etc.	50	70	94	96	(3–15 grams, 2 of 3)
T. and C.	97	98	95	100	(3–12, 6–15 grams. Both correct, but 2nd trial allowed)
T., T., W.	60	88	100	100	(3–12. 6–15 grams, 2 of 3)

V[2] Naming Four Colors

	4	5	6	7
T., L., O., etc.	59	74	86	97
T. and C.	65	74	84	96
T., T., W.	50	88	95	97

V[3] Aesthetic Comparison

	4	5	6	7
T., L., O., etc.	64	73	94	96
T. and C.	80	88	83	90
T., T., W.	60	86	90	98

V[4] Definitions—Use or Better

	4	5	6	7	
T., L., O., etc.	51	69	92	98	(Ball, fork, table, chair, horse, pencil, 4 of 6)
T. and C.	86	91	83	90	(Fork, table, chair, horse, mamma, 3 of 5)
T., T., W.	60	100	100	100	(Like Terman and Childs)

V[5] Patience, or Divided Rectangle

	4	5	6	7	
T., L., O., etc.	21	70	92	95	(2 of 3. 1 minute each)
T. and C.	70	88	95	100	(2 of 3. No time limit)
T., T., W.	50	75	90		(2 of 3. 1 minute each)

V[6] Three Commissions

	3	4	5	6	7	
T., L., O., etc.	20	50	72	91	93	(Order must be correct)
T. and C.		72	81	87	95	(Correct order not required)
T., T., W.		50	80	94	95	(Correct order)

V Alternative. Giving Age

	3	4	5	6	7
T., L., O., etc.	0	50	76	92	98
T. and C.		72	79	96	95

VI[1] Right and Left

	4	5	6	7	8	
T., L., O., etc.	40	50	71	86	95	(3 of 3, or 5 of 6)
T. and C.	41	73	75	90	93	(Like Binet)
T., T., and W.	25	63	62	72	90	(Like Binet)

VI[2] Mutilated Pictures

	4	5	6	7	8
T., L., O., etc.	27	50	65	87	96
T. and C.	34	48	57	68	82
T., T., and W.	12	33	74	86	97

VI[3] Counting Thirteen Pennies

	4	5	6	7	8
T., L., O., etc.	30	46	76	93	96
T. and C.		55	74	100	100
T., T., and W.		25	67	92	94

VI[4] Comprehension, 2nd Degree

	4	5	6	7	8
T., L., O., etc.	25	55	70	86	93

VI[5] Naming Four Coins

	4	5	6	7	8
T., L., O., etc.	25	47	74	91	95
T. and C.	19	41	69	93	100*
T., T., and W.		25	81	95	98

* Re-scored by new standard. Figures in original article allowed no error.

VI[6] Repeating 16–18 Syllables

	4	5	6	7	8	
T., L., O., etc.	34	56	69	90	95	(1 of 3 correct, or 2 with 1 error each)
T. and C.	31	53	62	80		(1 of 3 without error)
T., T., and W.		38	67	77	86	(1 of 3 without error)

VI Alternative. Forenoon and Afternoon

	4	5	6	7
T., L., O., etc.	46	60	82	97
T. and C.	65	84	87	100
T., T., and W.	51	68	88	95

VII[1] Number of Fingers

	5	6	7	8	9
T., L., O., etc.	24	51	72	86	95
T. and C.	33	73	92	97	100
T., T., and W.	25	57	80	88	98

VII[2] Pictures—Descriptions

	5	6	7	8	9	
T., L., O., etc.	27	56	63	88	97	
T. and C.	42	47	60	70	83	(Re-scored since 1912)
T., T., and W.	23	66	77	95	100	

VII[3] Repeating Five Digits

	5	6	7	8	9
T., L., O., etc.	34	59	74	83	93
T. and C.	50	50	72	74	83
T., T., and W.	26	43	74	80	88

VII[4] Tying Bow-knot

	5	6	7	8	9	
T., L., O., etc.	11	35	69	88	94	
T. and C.		27	61	74		(160 school children, ages 6, 7, 8)

VII[5] Giving Differences

	5	6	7	8	9	
T., L., O., etc.	23	54	66	78	90	
T. and C.	19	47	62	74	85	(Re-scored since 1912)
T., T., and W.	30	61	74	86	91	

VII[6] Copying Diamond

	5	6	7	8	9	
T., L., O., etc.	4	30	64	83	94	(Pen, 2 of 5)
T. and C.	16	42	67	90	98	(Pencil, 1 trial. Re-scored since 1912)
T., T., and W.	10	48	70	82	91	(Pencil, 1 trial)

VII Alternative 1. Days of Week

	5	6	7	8	9	
T., L., O., etc.	0	27	65	81	91	(No error. 2 of 3 "checks")
T. and C.		56	85	85	98	(1 error allowed. No "checks" used)

VII Alternative 2. Three Digits Backwards

	5	6	7	8	9	10
T., L., O., etc.	2	35	60	83	90	96

VIII[1] Ball and Field (Inferior Plan)

	6	7	8	9	10	11	12	13	14
T., L., O., etc.	37	52	60	67	73	77	82	87	90

VIII[2] Counting 20–0

	6	7	8	9	10
T., L., O., etc.	16	48	66	81	96
T. and C.	7	62	69	95	100
T., T., and W.	19	38	57	82	97

VIII[3] Comprehension, Third Degree

	6	7	8	9	10
T., L., O., etc.	47	59	72	85	92
T. and C.	38	57	69	78	85
T., T., and W.	20	51	62	72	81

VIII[4] Giving Similarities

	6	7	8	9	10	11
T., L., O., etc.	30	51	63	78	90	92

VIII[5] Definitions Superior to Use

	7	8	9	10	
T., L., O., etc.	43	62	71	83	(Balloon, tiger, football, soldier, 2 of 4)
T. and C.	34	58	77	89	(Fork, spoon, chair, horse, mamma. 2 of 4. Re-scored since 1912)

VIII[6] Vocabulary, 20 Definitions

	6	7	8	9	10	11
T., L., O., etc.	0	12	56	78	97	
T. and C.	0	14	61	69	91	
T., T., and W.	20	24	57	86	88	
Hopwood*				84	90	96
Houser*			63	80	93	98

* The tests made by Miss Margaret Hopwood and J. D. Houser, graduate students at Stanford University, were mass tests (written) on unselected children in the grades. 25 to 35 were tested at each age indicated.

VIII ALTERNATIVE 1. NAMING 6 COINS

	6	7	8	9	10	
T., L., O., etc.	27	40	64	88	95	
T. and C.	0	33	60	81	90	(Re-scored since 1912)

VIII ALTERNATIVE 2. WRITING FROM DICTATION

	6	7	8	9	10	
T., L., O., etc.	12	53	88	96	100	(Time limit, 1 minute)
T. and C.	7	79	83	100	100	(No time limit)

IX[1] GIVING DATE

	7	8	9	10	11	12	
T., L., O., etc.	20	50	67	83	91		
T. and C.	25	52	81	77	87	93	
T., T., and W.	14	48	71	86	94		
Knollin's unemployed				65	80	98	(By mental age)

IX[2] ARRANGING WEIGHTS

	7	8	9	10	11	12	
T., L., O., etc.	7	35	58	69	75	79	(2 of 3 trials correct)
T. and C.	25	29	55	54	77		(Only 1 trial given)
T., T., and W.		50	71	73	82		(2 of 3 trials correct)
Williams' delinquents				77	86	93	(By mental age)
Knollin's unemployed					70	79	(By mental age)

IX[3] MAKING CHANGE

	7	8	9	10	11	12	
T., L., O., etc.	3	38	60	83	92		(10–6, 15–12, 25–4, 2 of 3)
T. and C.	7	26	42	55	67		(25–9. One trial)
T., T., and W.	0	29	62	78	91		(25–4. One trial)
W's delinquents (By mental age)				85	100	98	
K's unemployed (By mental age)				95	98	100	(Note influence of age)

IX[4] REPEATING FOUR DIGITS BACKWARDS

	7	8	9	10	11	12	
T., L., O., etc.	18	44	62	75	86	91	
W's delinquents			46	61	73	82	(By mental age)

IX[5] THREE WORDS

	8	9	10	11	12	13	
T., L., O., etc.	44	68	81	90	95	94	
T. and C.	42	60	84	86	93		
T., T., and W.	48	68	86	91	90	90	
W's delinquents				95	100	100	(By mental age)
K's unemployed			30	67	75	96	(By mental age)

IX[6] FINDING RHYMES

	8	9	10	11	12	13	14
T., L., O., etc.	48	62	81	83	94		
T. and C.		74	92	81	82	89	
W's delinquents (By mental age)			60	66	80	85	92
K's unemployed (By mental age)			23*	28	73	81	90

* Note poor record of adults in this test.

IX ALTERNATIVE 1. NAMING THE MONTHS

	8	9	10	11	12	13	14	
T., L., O., etc.	30	59	78	90	93	96		(2 of 3 checks correct)
T. and C.	64	91	81	96				(No check given)
W's delinquents (Mental age)			78	87	88	95		(2 of 3 checks)
K's unemployed (Mental age)			88	94	100	100		

IX ALTERNATIVE 2. STAMPS

	7	8	9	10	11	12
T., L., O., etc.	13	39	69	90	96	97
T. and C.	58	60	73	85	90	100
T., T., and W.	26	50	59	84	91	100

X[1] VOCABULARY, 30 DEFINITIONS

	8	9	10	11	12	13	
T., L., O., etc.	0	25	71	80	100	100	
T. and C.	14	35	60	72	77	90	
T., T., and W.	11	23	56	73	88	95	
Hopwood			65	82	90	93	(Mass tests, written)
Houser			61	79	86	90	(Mass tests, written)
W's delinquents				100	100	100	(By mental age)
K's unemployed				100	100	100	(By mental age)

Terman's high-school pupils, all passed
K's business men, all passed

X[2] ABSURDITIES

	8	9	10	11	12	13	
T., L., O., etc.	16	47	64	67	75	85	
T. and C.	30	40	52	55	73	88	
T., T., and W.	29	50	71	80	85	88	
W's delinquents			68	79	100	84	(By mental age)
K's unemployed			50	60	82	81	(By mental age)

High-school pupils, all passed

X[3] DRAWING DESIGNS

	8	9	10	11	12	13	14	
T., L., O., etc.	27	46	60	72	81			
T., T., and W.	36	50	61	72	90			
W's delinquents			67	75	88	96	88	(By mental age)
K's unemployed			41	52	67	70	73	(By mental age)*

* Note inferiority of adults as compared with school children of same mental age.

X^4 Reading for 8 Memories

	8	9	10	11	12	13	
T., L., O., etc.	26	55	69	80	93		
T. and C.	13	46	58	65	62		(Re-scored for 8 memories)
T., T., and W.	30	50	64	74	79		
W's delinquents			62	75	88		(By mental age)

X^5 Comprehension, 4th Degree

	9	10	11	12	13	14	
T., L., O., etc.	44	60	74	83	88	91	
T. and C.	37	59	70	80	86		
T., T., and W.	45	62	65	78	90		
W's delinquents		52	70	86	97	100	(By mental age)
K's unemployed		70	85	98	100	100	(By mental age)

X^6 Naming 60 Words

	8	9	10	11	12	13	14	
T., L., O., etc.	21	50	63	76	85			(60 in 3 minutes)
T. and C.	35	57	67	83	82			(50 required in 2 minutes)
T., T., and W.	32	51	62	69	90			(50 in 2 minutes)
W's delinquents (Mental age)			30	65	75	86		(60 in 3 minutes)
K's unemployed (Mental age)				40	60	56	60	(60 in 3 minutes*)

X Alternative 1. Six Digits

	8	9	10	11	12	13	14	
T., L., O., etc.	32	56	71	80	87			
W's delinquents			46	68	73	81	86	(By mental age)
K's unemployed			54	70	71	98	100	(By mental age)

X Alternative 2. 20–22 Syllables

	8	9	10	11	12	13	14	15	16	
T., L., O., etc.	35	52	63	76	82					
W's delinquents				45	58	55	82	86	95	(By mental age)
K's unemployed				40	53	74	85	100	92	(By mental age)

X Alternative 3. Healy-Fernald Construction Puzzle

	8	9	10	11	12	13	
T., L., O., etc.	31	46	67	81	90		
W's delinquents				75	88	92	(By mental age)
K's unemployed				80	86	98	(By mental age)

* Note great inferiority of adult subjects, by mental age. These did not seem to enter into the spirit of the test as children do.

XII[1] Vocabulary, 40 Definitions

	9	10	11	12	13	14	
T., L., O., etc.	2	14	58	65	78	81	
T. and C.	6	15	44	57	62		
T., T., and W.	7	19	35	60	79		
Hopwood		20	46	65	80		(Mass tests)
Houser	4	16	41	58	70		(Mass tests)
W's delinquents			10	35	62	78	(By mental age)
K's unemployed		20	55	66	93	100	(By mental age)
T's high-school pupils.	All passed.						
K's business men.	All passed.						
T's college students.	All passed.						(Mass tests of 65)

XII[2] Abstract Words

	10	11	12	13	14	
T., L., O., etc.	27	48	57	63	72	
T. and C.	20	38	51	65	67	(Binet's words. Re-scored since 1912)
W's delinquents		39	37	75	77	(By mental age)
K's unemployed		50	90	83	97	(By mental age)
T's high-school pupils	(91 percent of all passed)					
K's business men	(90 percent of all passed)					

XII[3] Ball and Field (Superior Plan)

	8	9	10	11	12	13	14	15	16	
Combined Stanford data for 908 cases	17	25	38	52	60	67	72			
W's delinquents			54	65	81	78	86	84	81	(Mental age)
K's unemployed			51	58	60	66	74	90	85	(Mental age)
T's high-school pupils	(80 percent passed)									
K's business men.	(83 percent passed)									

XII[4] Dissected Sentences

	10	11	12	13	14	
T., L., O., etc.	25	49	62	71	82	
T. and C.	24	32	62	66	83	
T., T., and W.	37	50	57	76	80	
W's delinquents		21	55	66	77	(By mental age)
K's unemployed		0	27	54	63	(By mental age)
T's high-school pupils.	(95 percent passed)					
K's business men.	(85 percent passed)					

XII[5] Interpretation of Fables, Score 4

	10	11	12	13	14	15	16	
T., L., O., etc	37	53	62	70	81			
T. and C.	44	58	67	70	82			(Re-scored since 1912)
W's delinquents		30	53	64	68	83	100	(By mental age)
K's unemployed		20	28	52	60	73	94	(By mental age)
T's high-school pupils.	(90 percent passed)							
K's business men.	(83 percent passed)							

XII[6] FIVE DIGITS BACKWARDS

	10	11	12	13	14	15	16	
T., L., O., etc.	40	61	66	71	73			
W's delinquents		30	45	56	68	80	95	(By mental age)
K's unemployed		33	48	55	60	67	78*	(By mental age)
T's high-school pupils	(75 percent passed)							
K's business men.	(79 percent passed)							

* Note inferiority of older subjects as compared with children of same mental age.

XII[7] PICTURES, INTERPRETATION

	10	11	12	13	14	15	
T., L., O., etc.	42	50	63	74	80		
T., T., and W.	30	47	65	78	84		
W's delinquents		35	46	54	62	71	(By mental age)
K's unemployed		50	67	71	87	82	(By mental age)
T's high-school pupils.	(95 percent passed)						
K's business men.	(93 percent passed)						

XII[8] FINDING SIMILARITIES, THREE THINGS

	10	11	12	13	14	15	16	
T., L., O., etc.	43	57	64	73	78			
W's delinquents		44	60	84	91	96		(By mental age)
K's unemployed		50	52	56	63	73	75	(By mental age)
T's high-school pupils.	(95 percent passed)							
K's business men.	(76 percent passed)							

XIV[1] VOCABULARY, 50 DEFINITIONS

	10	11	12	13	14	15	16	
T., L., O., etc.	0	0	20	45	66			
T. and C.	4	18	25	42	50			
T., T., and W.	3	17	23	35				
Hopwood	0	6	32	40	67			(Mass tests)
Houser	0	10	24	51	60			(Mass tests)
W's delinquents	0	5	10	40	58	70	90	(By mental age)
K's unemployed	10	35	46	73	80	90	100	(By mental age)*
T's high-school pupils.	(All passed)							
K's business men.	(All passed)							
T's college students.	(All passed)							

* Note superiority of adults as compared with children of the same mental age.

XIV[2] INDUCTION TEST; FINDING A RULE

	11	12	13	14	15	16	
T., L., O., etc.	17	38	47	56			
W's delinquents	14	45	72	79	86	85	(By mental age)
K's unemployed	39	44	71	80	100	100	(By mental age)
T's high-school pupils.	(78 percent passed)						
K's business men.	(90 percent passed)						

XIV[3] President and King

	11	12	13	14	15	16	
T., L., O., etc.	10	22	39	52			
T., T., and W.	18	34	44	55			
W's delinquents	13	20	34	59	68	76	(By mental age)
K's unemployed	20	28	46	67	64	87	(By mental age)
T's high-school pupils.	(88 percent passed)						
K's business men.	(80 percent passed)						

XIV[4] Problems of Fact

	11	12	13	14	15	16	
T., L., O., etc.	31	39	49	56			
T. and C.	40	41	48	100			
T., T., and W.	38	50	61	70			
W's delinquents	21	36	53	64	77	86	(By mental age)
K's unemployed	40	70	82	85	92	94	(By mental age)
T's high-school pupils.	(90 per cent passed)						
K's business men.	(100 percent passed)						

XIV[5] Arithmetical Reasoning

	11	12	13	14	15	16	
T., L., O., etc.	26	38	43	50			
T., T., and W.	10	30	41	49			
W's delinquents	0	22	34	45	62	86	(By mental age)
K's unemployed	10	28	48	73	82	87	(By mental age)
T's high-school pupils.	(81 percent passed)						
K's business men.	(79 percent passed)						

XIV[6] Reversing Hands of Clock

	11	12	13	14	15	
T., L., O., etc.	30	54	64	78		
T., T., and W.	20	59	77	89		
W's delinquents	17	48	75	81	86	(By mental age)
K's unemployed	10	50	63	83	91	(By mental age)
T's high-school pupils.	(90 percent passed)					
K's business men.	(86 percent passed)					

XIV Alternate. Repeating Seven Digits

	11	12	13	14	15	16	
T., L., O., etc.	36	41	52	59			(1 of 2 correct)
T. and C.	43	48	57	60			(Re-scored for 1 of 2 correct)
T., T., and W.	37	46	56	65			
W's delinquents	27	30	36	52	63	78	(By mental age)
K's unemployed	10	22	44	50	65	82	(By mental age)*
T's high-school pupils.	(60 percent passed)						
K's business men.	(61 percent passed)						

* Note inferiority of adults as compared with children of same mental age.

AVERAGE ADULT 1. VOCABULARY, 65 DEFINITIONS

	12	13	14	15	16	
T., L., O., etc.	0	12	23			
T. and C.	4	9	22			
T., T., and W.	0	8	14			
Hopwood	0	6	13			(Mass tests)
Houser	3	9	28			(Mass tests)
W's delinquents	0	0	4	32	62	(By mental age)
K's unemployed	0	11	24	53	77	(By mental age)
T's high-school pupils.	(84 percent passed)					
K's business men.	(75 percent passed)					
T's college students.	(96 percent passed)					

AVERAGE ADULT 2. INTERPRETATION OF FABLE, SCORE 8

	12	13	14	15	16	
T., L., O., etc.	12	23	31			
T. and C.	16	27	36			
W's delinquents	0	12	40	47	60	(By mental age)
K's unemployed	6	8	20	39	65	(By mental age)
T's high-school pupils.	(66 percent passed)					
K's business men.	(60 percent passed)					

AVERAGE ADULT 3. DIFFERENCES BETWEEN ABSTRACT WORDS

	11	12	13	14	15	16	
T., L., O., etc.		11	26	39			
T. and C.	16	24	29	44			
W's delinquents	0	4	12	18	46	60	(By mental age)
K's unemployed	0	0	25	43	55	76	(By mental age)*
T's high-school pupils.	(53 percent passed)						
K's business men.	(66 percent passed)						

* Note inferiority of adults as compared with normal children of same mental age.

AVERAGE ADULT 4. PROBLEM OF ENCLOSED BOXES

	12	13	14	15	16	
T., L., O., etc.	8	12	22			
W's delinquents	24	30	38	52	70	(By mental age)
K's unemployed	10	13	40	38	70	(By mental age)
T's high-school pupils.	(55 percent passed)					
K's business men.	(65 percent passed)					

AVERAGE ADULT 5. REPEATING OF SIX DIGITS BACKWARDS

	12	13	14	15	16	17	
T., L., O., etc.	10	32	45				
W's delinquents	4	17	30	50	70		(By mental age)
K's unemployed	6	14	24	30	57	69	(By mental age)
T's high-school pupils.	(56 percent passed)						
K's business men.	(60 percent passed)						

AVERAGE ADULT 6. LEARNING A CODE

	12	13	14	15	16	17	
T., L., O., etc.	8	17	32				
T., T., and W.	10	19	26				
W's delinquents	10	20	50	62	70		(By mental age)
K's unemployed	0	0	0	0	19	73	(By mental age)
T's high-school pupils.	(60 percent passed)						
K's business men.	(42 percent passed)						

AVERAGE ADULT. ALTERNATIVE 1. 28 SYLLABLES

	11	12	13	14	15	16	
T., L., O., etc.	10	26	35	39			
T. and C.	21	33	42	50			
T., T., and W.	15	37	40	48			
W's delinquents	8	14	24	35	46	58	(By mental age)
K's unemployed	0	12	20	31	42	57	(By mental age)
T's high-school pupils.	(44 percent passed)						
K's business men.	(59 percent passed)						

AVERAGE ADULT 2. COMPREHENSION OF PHYSICAL RELATIONS

	12	13	14	15	16	
T., L., O., etc.	5	19	33			
W's delinquents	31	38	46	53	60	(By mental age)
K's unemployed	10	22	59	46	74	(By mental age)*
T's high-school pupils.	(65 percent passed)					
K's business men.	(80 percent passed)					

* Note superiority of older subjects as compared with children of same mental age.

SUPERIOR ADULT 1. VOCABULARY, 75 DEFINITIONS

	15	16	17	18	
W's delinquents	5	10	40		(By mental age)
K's unemployed	5	38	60		(By mental age)
T's high-school pupils.	(61 percent passed)				
K's business men.	(50 percent passed)				
T's college students.	(85 percent passed)				

SUPERIOR ADULT 2. BINET'S PAPER-CUTTING TEST

	12	13	14	15	16	17	18	
T., L., O.	2	13	25					
K's unemployed	0	0	11	27	38	54	61	(By mental age)
T's high-school pupils.	(30 percent passed)							
K's business men.	(43 percent passed)							

SUPERIOR ADULT 3. REPEATING EIGHT DIGITS

	12	13	14	15	16	17	18	
K's unemployed	6	9	11	27	34	53	62	(By mental age)
T's high-school pupils.	(37 percent passed)							
K's business men.	(44 percent passed)							

SUPERIOR ADULT 4. SENSE OF RELATION

	12	13	14	15	16	17	
T., L., O., etc.....	13	27	30				
W's delinquents.....	0	16	20	24			(By mental age)
K's unemployed.....	0	9	18	25	38	50	(By mental age)
T's high-school pupils.	(65 percent passed)						
K's business men.	(26 percent passed)						

SUPERIOR ADULT 5. SEVEN DIGITS BACKWARDS

T's high-school pupils. (40 percent passed)
T's 14-yr.-olds in 8th grade. (13 percent passed)

SUPERIOR ADULT 6. INGENUITY TEST

	13	14	15	16	17	18	
T., L., O., etc.....	6	11					
W's delinquents.....	8	6	30	43			(By mental age)
K's unemployed.....	0	14	16	29	47	57	(By mental age)
T's high-school pupils.	(25 percent passed)						
K's business men.	(36 percent passed)						

APPENDIX II

Form Used for Gathering Supplementary Information

To the Teacher

The information rendered on this blank will be held strictly confidential, will be used for statistical purposes only, and without the mention of any child's name. The information will aid greatly in connection with a revision of the Binet mental tests.

Even if the child's social status is not accurately known, please underscore the word which you *think* would properly represent it, and if the uncertainty is very great, add a question mark.

"Additional information," as called for in the last item, will be especially welcome.

1. Name of pupil..... Age.....
2. Name of school.....
3. Present grade..... (Grade attended March, 1914, was.....)
4. Quality of child's school work (underscore appropriate word): Very inferior, inferior, average, superior, very superior.
5. Teacher's estimate of child's intelligence as compared with average children of the same age (underscore): Very inferior, inferior, average, superior, very superior.
6. Social class to which the child belongs (this refers to the intellectual level, the culture and the general level of the home environment): Very inferior, inferior, average, superior, very superior.
7. Additional information which will throw light on the child's intelligence, school success or social status.

Psych
T319st

599981

Terman, Lewis Madison
The Stanford revision and extension of the Binet-Simon scale for measuring intelligence

University of Toronto
Library

DO NOT
REMOVE
THE
CARD
FROM
THIS
POCKET